A Generalization to Schur's Lemma with an Application to Joint Independent Subspace Analysis

Dana Lahat, Christian Jutten

A Generalization to Schur's Lemma
with an Application to Joint Independent Subspace Analysis

Dana Lahat and Christian Jutten, *Fellow, IEEE*[*†]

Abstract

This paper has a threefold contribution. First, it introduces a generalization to Schur's lemma from 1905 on irreducible representations. Second, it provides a comprehensive uniqueness analysis to a recently-introduced source separation model. Third, it reinforces the link between signal processing and representation theory, a field of algebra that is more often associated with quantum mechanics than signal processing. The source separation model that this paper relies on performs joint independent subspace analysis (JISA) using second order statistics. In previous work, we derived the Fisher information matrix (FIM) that corresponds to this model. The uniqueness analysis in this paper is based on analysing the FIM, where the core of the derivation is based on our proposed generalization to Schur's lemma. We provide proof both to the new lemma and to the uniqueness conditions. From a different perspective, the generalization to Schur's lemma is inspired by a coupled matrix block diagonalization problem that arises from the JISA model. The results in this paper generalize previous results about identifiability of independent vector analysis (IVA). This paper complements previously-known results on the uniqueness of joint block diagonalization (JBD) and block term decompositions (BTD), as well as of their coupled counterparts.

Preliminaries

Keywords Schur's lemma, irreducible representations, coupled decompositions, joint block diagonalization, uniqueness, identifiability, blind source separation, independent subspace analysis, independent vector analysis, data fusion

1 Introduction

A well-known result in algebra, group theory and irreducible representations ([1, 2, 3, 4] and others) is often known as "Schur's lemma" [5]. In previous work [6], Schur's lemma arose naturally in the analysis of the uniqueness of blind source separation (BSS) of a mixture of piecewise stationary real multidimensional sources. This BSS model can be reformulated as a symmetric joint block diagonalization (JBD) of a set of covariance matrices. The analysis in [6] boiled down to showing that the model is non-identifiable for JBD-irreducible data if and only if (iff) at least two multidimensional sources exist in equivalent subspaces. This result is complementary to the generic uniqueness analysis in [7]. In [7], it was shown that JBD was a special case of a more general family of tensor factorizations, called block term decomposition (BTD). In that paper, generic uniqueness conditions for BTD were derived, and it was stated that "In the nongeneric case, lack of uniqueness can be due to the fact that tensors can be further block-diagonalized", or "be subdivided in smaller blocks". Our result in [6] characterises the cases where symmetric JBD of covariance matrices, which is a special case of BTD, is non-identifiable in the nongeneric case and when the blocks are irreducible, that is, cannot be further divided into smaller blocks.

The analysis in [6] is based on characterising the non-invertibility of the Fisher information matrix (FIM). In this paper, we follow a similar analytical approach. Based on previously-derived FIM for a different BSS model, called joint independent subspace analysis (JISA) [8, 9], we characterise, in this paper, the singular points of this FIM. Interestingly, this procedure resulted in a different lemma on irreducible subspaces. This new lemma, which we present in this paper, can be regarded as a generalization of Schur's original lemma. In analogy to the results in [6], also in this case the non-identifiability conditions for the irreducible case (i.e., nongeneric uniqueness) can be stated as an equivalence of subspaces of a pair of sources. However, in this case, the equivalence is in a generalized sense, as we define in this paper.

The results in [6], which now can be regarded as a special case of those in this current paper, conform with those in [7] (and references therein) about invariant subspaces and uniqueness of BTD.

[*]D. Lahat and Ch. Jutten are with GIPSA-Lab, UMR CNRS 5216, Grenoble Campus, BP46, F-38402 Saint Martin d'Hères, France. email:{Dana.Lahat, Christian.Jutten}@gipsa-lab.grenoble-inp.fr.

[†]This work is supported by the project CHESS, 2012-ERC-AdG-320684. GIPSA-Lab is a partner of the LabEx PERSYVAL-Lab (ANR–11-LABX-0025).

JISA is a generalization of independent vector analysis (IVA) to multidimensional components, or subspaces. Accordingly, the uniqueness results in this paper generalize those in [10, 11, 12], for IVA, to the multidimensional case. This is in analogy to the results in [7, 6], which generalize the well-known "spectral diversity" condition for one-dimensional BSS with temporal or spectral diversity, to multidimensional subspaces.

The interpretation of the results in this paper, in terms of signal processing and data fusion, follows similar lines as its IVA counterpart in [10, 11, 12] and references therein.

We mention that in the IVA or non-stationary one-dimensional BSS case there is no need to resort to Schur's lemma (or a variant thereof), because the subspaces are one-dimensional and the corresponding "blocks" are scalar; hence, there is no irreducibility issue: the blocks in the algebraic formulation of the model, be it joint diagonalization (JD) or coupled JD [13, 14], are already of size one, and thus cannot be further reduced. For this reason, the derivation of uniqueness of IVA in [10] does not involve Schur's lemma despite the fact that it follows the same methodology as in this paper.

In this paper, we chose, for simplicity and clarity of exposition, to deal with a simple instance of JISA. However, more elaborate formulations exist (e.g. [15],[16, Section VI]) that allow variability between data sets, for example in the size of the blocks, as well as mixing matrices of different sizes, possibly rectangular. We assume real data but our results can be readily generalized to the complex domain. In particular, one can join the non-stationarity diversity of [6] with the multiset diversity of JISA, as a natural generalization of the model in e.g. [13, 14]. The model in [13, 14] amounts to a coupled JD, whereas in our case, the model would amount to coupled JBD [9]. One can combine both results of uniqueness, [6] and of this paper, and conclude that for coupled JBD, the model is not identifiable if one can find equivalence of subspaces both in the data set and in the temporal or frequency coordinates. In particular, JBD/BTD has strong uniqueness also in the underdetermined case [7]. As a result, these properties are inherited by the coupled formulation and further reinforce its uniqueness. Roughly speaking, the larger the number of types of diversity in the model, the smaller the risk to fall randomly into a situation of non-uniqueness.

Our interest in the case of second-order statistics (SOS)-based real-valued JISA with only data set diversity and no spatio/temporal diversity (as in, e.g., [13, 14, 17]) stems from the fact that this is the only setup in which there is absolutely no identifiability of each data set alone. Hence, it is the most challenging scenario, and illustrates the power of SOS-based JISA. The observation that coupled factorizations may be unique even when individual factorizations are not unique, is not new, see, e.g., [18, 11, 12, 19] and references therein.

The link between irreducible representations and JBD in the context of signal processing has first been introduced, to the best of our knowledge, by [20, 21, 22, 23]. The main difference of this work from these (and from [6]) is that in this work we provide a substantial generalization to well-known results in algebra, and this is achieved by analysing a new type of a BSS model as well as a new type of coupled factorization.

The main novelty and contribution of this paper is the lemmas in Section 3 and the JISA identifiability Theorem 4.1. Some of the results on partitioned matrices, in Appendix A, are also new.

Part of this work was presented in [24].

The notations in this paper follow those of [9], unless stated otherwise. $\mathbb{K}$ denotes real or complex numbers. Accordingly, $\cdot^{\dagger}$ denotes transpose or conjugate transpose. $\cdot^{\top}$ is the ordinary transpose. $\mathbf{A}^{[k]} \triangleq (\mathbf{A}^{[k]})^{-1}$.

The rest of this paper is as follows. In Section 2 we introduce some properties of sets of matrices that will be used throughout this paper. Section 3 presents our generalization to Schur's lemma. In Section 4, we apply the new lemma to the uniqueness analysis of a basic form of JISA. We briefly discuss our results in Section 5.

2 Basic Definitions

The following definitions will be used throughout this paper.

2.1 Equivalence, Irreducibility

Definition 2.1 (Irreducibility (in the generalized sense))**.** Consider a set of matrices $\mathbf{P}^{[k,l]} \in \mathbb{K}^{P^{[k]} \times P^{[l]}}$, $k, l = 1, \ldots, K$. This set is reducible in the generalized sense if there exist K invertible matrices (transformations) $\mathbf{T}^{[k]} \in \mathbb{K}^{P^{[k]} \times P^{[k]}}$ such that

$$\mathbf{T}^{[k]}\mathbf{P}^{[k,l]}(\mathbf{T}^{[l]})^{\dagger} \quad \in \mathbb{K}^{P^{[k]} \times P^{[l]}} \quad \forall k, l$$

can all be brought to a form $\begin{bmatrix} \mathbf{P}_1^{[k,l]} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2^{[k,l]} \end{bmatrix}$, where $\mathbf{P}_1^{[k,l]}$ and $\mathbf{P}_2^{[k,l]}$ are rectangular matrices of size $P_i^{[k]} \times P_i^{[l]}$, $i = 1, 2$, and $P_i^{[k]} \geq 1 \ \forall i, k$. Otherwise, the set is irreducible (in the generalized sense).

Remark 1. When $K \leq 2$, under certain conditions, the set $\{\mathbf{P}^{[k,l]}\}$ may be exactly diagonalized, for example, using generalized eigenvalue decomposition (GEVD) [25, Chapter 12.2, Equation (53)]. In this case, irreducibility applies only to $K \geq 3$.

Definition 2.2 (Similarity and equivalence (in the generalized sense))**.** Two sets of matrices, $\{\mathbf{R}^{[k,l]}\}_{k,l=1}^{K}$ and $\{\mathbf{P}^{[k,l]}\}_{k,l=1}^{K}$, of size $P^{[k]} \times P^{[l]}$, are similar in the generalized sense if they are related by a generalized similarity transformation

$$\mathbf{P}^{[k,l]} = \mathbf{\Psi}^{[k]}\mathbf{R}^{[k,l]}\mathbf{\Psi}^{-[l]}\ , \quad \forall k,l$$

for some K invertible matrices $\{\mathbf{\Psi}^{[k]}\}_{k=1}^{K}$ of size $P^{[k]} \times P^{[k]}$. Generalized similarity can be regarded as a generalized equivalence relation.

2.2 Admissible Data

The new lemma is inspired by a signal processing model that imposes certain constraints on the data. These constraints correspond to certain plausible physical assumptions. We now define types of data admissible by our lemma. First, we make a distinction between two types of admissible data: "strict", and "relaxed". The "strict" form corresponds to setups more similar to those used in Schur's original lemma or its immediate variants. "Relaxed" admits a broader range of scenarios that are plausible in BSS, JISA and data fusion. Within the JISA framework, "strict" corresponds to sources that are mutually dependent across all mixtures, whereas "relaxed", as its name implies, relaxes this assumption. Second, the lemma admits two different types of matrices. "Symmetric" corresponds to the JISA model [9], in which each matrix reflects (cross-) covariance between true multidimensional components. The "unitary" scenario emphasizes the analogy between this lemma and Schur's original; we use it only in the "strict" sense.

Definition 2.3 (Admissible and non-admissible data)**.** A set of matrices $\mathbf{P}^{[k,l]} \in \mathbb{K}^{P^{[k]} \times P^{[l]}}$, $k,l = 1,\ldots,K$ is admissible if any of the following holds:

Unitary: all $\mathbf{P}^{[k,l]}$ unitary (implies $P^{[k]} = P\ \forall k$) and the set $\{\mathbf{P}^{[k,l]}\}_{k,l=1}^{K}$ irreducible

Symmetric: $\mathbf{P}^{[k,l]}$ is the (k,l)th block of $\mathbf{P}$, a matrix of size $\left(\sum_{k=1}^{K} P^{[k]}\right) \times \left(\sum_{k=1}^{K} P^{[k]}\right)$, where

1. $\mathbf{P}$ has a non-zero determinant
2. $\mathbf{P}$ Hermitian, which implies $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^{\dagger}\ \forall k,l$

and either

Strict:
1. $\mathbf{P}^{[k,l]}$ full-rank $\forall k,l$
2. the set $\{\mathbf{P}^{[k,l]}\}_{k,l=1}^{K}$ irreducible

or

Relaxed:
1. For each $k \neq l$, either $\mathbf{P}^{[k,l]}$ full-rank with no zero entries, or $\mathbf{P}^{[k,l]} = \mathbf{0}_{P^{[k]} \times P^{[l]}}$
2. If for some $k \neq l$ we have $\mathbf{P}^{[k,l]} = \mathbf{0}$, and $\mathbf{P}$ can be permuted to a block-diagonal form, i.e., can be rewritten as a direct sum of $B \geq 2$ smaller matrices $\mathbf{P} = \bigoplus_{b=1}^{B} \mathbf{P}^{[\mathcal{B}_b,\mathcal{B}_b]}$, where $\mathcal{B}_b$ is the bth subset in a partition of $\{1,\ldots,K\}$, then, for any $|\mathcal{B}_b| \geq 2$, the set of matrices in $\mathbf{P}^{[\mathcal{B}_b,\mathcal{B}_b]}$ must be admissible in the strict sense.

Otherwise, the set is not admissible.

Remark 2. When the data are symmetric, we may consider only $k \leq l$.

3 A Multiset Analogue to Schur's Lemma

3.1 A Multiset Analogue to Schur's First Lemma

We now present a multiset analogue to Schur's first lemma [5].

Lemma 1 (Multiset analogue to Schur's first lemma, strict)**.** *Let $\{\mathbf{P}^{[k,l]}\}$, $k,l = 1,\ldots,K$, be a set of full-rank $P^{[k]} \times P^{[l]}$ matrices, real or complex-valued, either symmetric in the sense $(\mathbf{P}^{[k,l]}) = (\mathbf{P}^{[l,k]})^{\dagger}$ or all unitary, i.e., admissible in the strict sense of Definition 2.3. If there exist K matrices $\mathbf{M}^{[1]},\ldots,\mathbf{M}^{[K]}$ of size $P^{[k]} \times P^{[k]}$ such that*

$$\mathbf{M}^{[k]}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{M}^{[l]} \quad \forall k,l$$

then $\mathbf{M}^{[k]} = \mu\mathbf{I}_{P^{[k]}}\ \forall k$, $\mu \in \mathbb{C}$.

Corollary 3.1 (Schur's first lemma)**.** *Schur's first lemma is a special case of Lemma 1, when $K = 1$.*

For symmetric data (Definition 2.3), we may relax some of the constraints. In this case, Lemma 1 rewrites as

Lemma 2 (Multiset analogue to Schur's first lemma, relaxed)**.** *Let $\{\mathbf{P}^{[k,l]}\}$, $k, l = 1, \dots, K$, be a set of $P^{[k]} \times P^{[l]}$ matrices, real or complex-valued, admissible in the relaxed sense of Definition 2.3. If there exist K matrices $\mathbf{M}^{[1]}, \dots, \mathbf{M}^{[K]}$ of size $P^{[k]} \times P^{[k]}$ such that*

$$\mathbf{M}^{[k]}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{M}^{[l]} \quad \forall k, l$$

then, either

1. $\mathbf{M}^{[k]} = \mathbf{0}\ \forall k$, *or*
2. *there exist, without loss of generality (w.l.o.g.), $\mathcal{D} = \{1, \dots, D\}$, $D \geq 2$, such that*

$$\mathbf{P}^{[k\in\mathcal{D}, l\notin\mathcal{D}]} = \mathbf{0} = (\mathbf{P}^{[k\notin\mathcal{D}, l\in\mathcal{D}]})^\dagger \quad \Leftrightarrow \quad \mathbf{P} = \left[\begin{array}{c|c} \mathbf{P}^{[1:D,1:D]} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}^{[D+1:K,D+1:K]} \end{array}\right]$$

   and $\mathbf{M}^{[k\in\mathcal{D}]} = \mu\mathbf{I}_{P^{[k]}}$, $\mu \in \mathbb{C}$, or
3. *there exist, w.l.o.g., $\mathcal{D} = \{1, \dots, D\}$, $D \geq 1$, such that*

$$\mathbf{P}^{[k\in\mathcal{D}, l\neq k]} = \mathbf{0} = (\mathbf{P}^{[k\neq l, l\in\mathcal{D}]})^\dagger \quad \Leftrightarrow \quad \mathbf{P} = \left[\begin{array}{ccc|c} \mathbf{P}^{[1,1]} & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \mathbf{P}^{[D,D]} & \\ \hline & \mathbf{0} & & \mathbf{P}^{[D+1:K,D+1:K]} \end{array}\right]$$

   and $\mathbf{M}^{[k\in\mathcal{D}]} = \mathbf{U}^{[k]}\operatorname{diag}\{\mu_1^{[k]}, \dots, \mu_{P^{[k]}}^{[k]}\}(\mathbf{U}^{[k]})^\dagger$, $\mu_p^{[k]} \in \mathbb{C}$, where $\mathbf{U}^{[k]}$ are the unitary matrices of the eigenvalue decomposition (EVD) of $\mathbf{P}^{[k,k]}$.

In the last two cases, without additional information about $\mathbf{P}^{[D+1:K,D+1:K]}$, we may set $\mathbf{M}^{[k\notin\mathcal{D}]} = \mathbf{0}$.

The proof of Lemma 1 and Lemma 2 is given in Appendix B.

3.2 A Multiset Analogue to Schur's Second Lemma

We now present a multiset analogue to Schur's second lemma [5].

Lemma 3 (Multiset analogue to Schur's second lemma, strict)**.** *Let $\mathcal{R}$, $\mathcal{P}$ be two irreducible (in the generalized sense of Definition 2.1) sets of invertible matrices $\mathbf{R}^{[k,l]}$, $\mathbf{P}^{[k,l]}$, $k, l = 1, \dots, K$, of size $R^{[k]} \times R^{[k]}$ and $P^{[k]} \times P^{[k]}$, respectively, either symmetric in the sense $\mathbf{R}^{[k,l]} = (\mathbf{R}^{[l,k]})^\dagger$, $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^\dagger$ or all unitary, i.e., admissible in the strict sense of Definition 2.3. If there exist K matrices $\mathbf{L}^{[1]}, \dots, \mathbf{L}^{[K]}$ of size $P^{[k]} \times R^{[k]}$ such that*

$$\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{L}^{[l]} \quad \forall k, l\,, \tag{1}$$

then, if $R^{[k]} = P^{[k]}\ \forall k$, either $\mathbf{L}^{[k]} = \mathbf{0}_{P^{[k]}\times P^{[k]}}\ \forall k$ or $\mathbf{L}^{[k]} = \nu\mathbf{O}^{[k]}\ \forall k$ (where $\mathbf{O}^{[k]}$ unitary and $\nu \in \mathbb{C}$) such that $\mathcal{R}$, $\mathcal{P}$ are related by a generalized similarity transformation (an equivalence relation) $\mathbf{P}^{[k,l]} = \mathbf{O}^{[k]}\mathbf{R}^{[k,l]}\mathbf{O}^{[l]\dagger}\ \forall k, l$. If $P^{[k]} \neq R^{[k]}$ for at least one k, $\mathbf{L}^{[k]} = \mathbf{0}_{P^{[k]}\times R^{[k]}}\ \forall k$.

Corollary 3.2 (Schur's second lemma)**.** *Schur's second lemma is a special case of Lemma 3, when $K = 1$.*

For symmetric data (Definition 2.3), we may relax some of the constraints. In this case, Lemma 3 rewrites as

Lemma 4 (Multiset analogue to Schur's second lemma, relaxed)**.** *Let $\mathcal{R}$, $\mathcal{P}$ be two sets of matrices $\mathbf{R}^{[k,l]} = (\mathbf{R}^{[l,k]})^\dagger$, $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^\dagger$, $k, l = 1, \dots, K$, of size $R^{[k]} \times R^{[k]}$ and $P^{[k]} \times P^{[k]}$, respectively, admissible in the relaxed sense of Definition 2.3. If there exist K matrices $\mathbf{L}^{[1]}, \dots, \mathbf{L}^{[K]}$ of size $P^{[k]} \times R^{[k]}$ such that*

$$\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{L}^{[l]} \quad \forall k, l$$

Then, either

1. $\mathbf{L}^{[k]} = \mathbf{0}_{P^{[k]}\times R^{[k]}}\ \forall k$, *or*

2. there exist, w.l.o.g., $k \in \mathcal{D}$, $\mathcal{D} = \{1, \ldots, D\}$, $D \geq 2$, such that

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{P}^{[1:D,1:D]} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}^{[D+1:K,D+1:K]} \end{array}\right] \quad , \quad \mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}^{[1:D,1:D]} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}^{[D+1:K,D+1:K]} \end{array}\right] ,$$

and $\mathbf{L}^{[k]} = \nu \mathbf{O}^{[k]}$, where $\mathbf{O}^{[k]}$ unitary and $\nu \in \mathbb{C}$ (implicitly, $P^{[k]} = R^{[k]}\ \forall k$), such that $\mathbf{R}^{[k \in \mathcal{D}, l \in \mathcal{D}]}$ and $\mathbf{P}^{[k \in \mathcal{D}, l \in \mathcal{D}]}$ are related by a generalized similarity transformation (an equivalence relation) $\mathbf{P}^{[k,l]} = \mathbf{O}^{[k]} \mathbf{R}^{[k,l]} \mathbf{O}^{[l]\dagger}$ for $k, l \in \mathcal{D}$, or

3. there exist, w.l.o.g., $k \in \mathcal{D}$, $\mathcal{D} = \{1, \ldots, D\}$, $D \geq 1$, such that

$$\mathbf{P} = \left[\begin{array}{ccc|c} \mathbf{P}^{[1,1]} & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \mathbf{P}^{[D,D]} & \\ \hline & \mathbf{0} & & \mathbf{P}^{[D+1:K,D+1:K]} \end{array}\right] \quad , \quad \mathbf{R} = \left[\begin{array}{ccc|c} \mathbf{R}^{[1,1]} & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \mathbf{R}^{[D,D]} & \\ \hline & \mathbf{0} & & \mathbf{R}^{[D+1:K,D+1:K]} \end{array}\right]$$

where, for $k \in \mathcal{D}$, there exist $1 \leq Q^{[k]} \leq \min(P^{[k]}, R^{[k]})$ such that

$$\mathbf{P}^{[k,k]} = \left[\begin{array}{c|c} \mathrm{diag}\{[\mathbf{P}^{[k,k]}]_{1:Q^{[k]},1:Q^{[k]}}\} & \mathbf{0} \\ \hline \mathbf{0} & [\mathbf{P}^{[k,k]}]_{Q^{[k]}+1:P^{[k]},Q^{[k]}+1:P^{[k]}} \end{array}\right]$$

$$\mathbf{R}^{[k,k]} = \left[\begin{array}{c|c} \mathrm{diag}\{[\mathbf{R}^{[k,k]}]_{1:Q^{[k]},1:Q^{[k]}}\} & \mathbf{0} \\ \hline \mathbf{0} & [\mathbf{R}^{[k,k]}]_{Q^{[k]}+1:R^{[k]},Q^{[k]}+1:R^{[k]}} \end{array}\right]$$

Let $\mathbf{\Lambda}^{[k]}$ denote a non-singular $Q^{[k]} \times Q^{[k]}$ diagonal matrix and $\mathbf{\Pi}_P^{[k]}$, $\mathbf{\Pi}_R^{[k]}$ two arbitrary $Q^{[k]} \times Q^{[k]}$ permutation matrices. Then,

$$\mathbf{L}^{[k]} = \left[\begin{array}{c} \mathbf{\Pi}_P^{[k]} \\ \hline \mathbf{0} \end{array}\right] \mathbf{\Lambda}^{[k]} \left[\begin{array}{c|c} \mathbf{\Pi}_R^{[k]} & \mathbf{0} \end{array}\right] \quad \forall k \in \mathcal{D}$$

In the last two cases, without additional information about $\mathbf{P}^{[D+1:K,D+1:K]}$, we may set $\mathbf{L}^{[k \notin \mathcal{D}]} = \mathbf{0}$.

The proof of Lemma 3 and Lemma 4 is given in Appendix C.

Example 1. *In the last scenario of Lemma 4, let $\mathbf{P}^{[k,k]} = \mathrm{diag}\{\mathbf{P}^{[k,k]}\}$ and $\mathbf{R}^{[k,k]} = \mathrm{diag}\{\mathbf{R}^{[k,k]}\}$ for some $k \in \mathcal{D}$. Assume w.l.o.g. that $R^{[k]} \leq P^{[k]}$ for this k. This implies $Q^{[k]} = R^{[k]}$. Then, it follows from Lemma 4 that (1) holds for any full-rank $\mathbf{L}^{[k]} = \mathbf{\Pi}_P^{[k]} \mathbf{\Lambda}^{[k]} \mathbf{\Pi}_R^{[k]}$.*

4 Uniqueness and Identifiability of JISA

We now turn to the second part of this paper, which is the identifiability of JISA [8, 9]. JISA is a source separation model that at its simplest form can be reformulated as a coupled matrix block diagoanlization.

As in JBD [26], and in some similarity to the identifiability analysis of IVA in [27, 12], it is possible to study the identifiability of the model through the properties of the FIM, whenever it is available in closed form. For each pair (i, j), it has been shown in [9] that the FIM corresponds to the symmetric positive semi-definite $2Km_im_j \times 2Km_im_j$ matrix

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} & \mathbf{I}_K \otimes \boldsymbol{\mathcal{T}}_{m_j,m_i} \\ \mathbf{I}_K \otimes \boldsymbol{\mathcal{T}}_{m_i,m_j} & \mathbf{S}_{ii} \boxplus \mathbf{S}_{jj}^{-1} \end{bmatrix} \tag{2}$$

where

$$\mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} = \begin{bmatrix} \mathbf{S}_{jj}^{[1,1]} \otimes [\mathbf{S}_{ii}^{-1}]_{11} & \cdots & \mathbf{S}_{jj}^{[1,K]} \otimes [\mathbf{S}_{ii}^{-1}]_{11} \\ \vdots & & \vdots \\ \mathbf{S}_{jj}^{[K,1]} \otimes [\mathbf{S}_{ii}^{-1}]_{K1} & \cdots & \mathbf{S}_{jj}^{[K,K]} \otimes [\mathbf{S}_{ii}^{-1}]_{KK} \end{bmatrix}$$

is a $Km_im_j \times Km_im_j$ matrix whose (k,l)th block is $\mathbf{S}_{jj}^{[k,l]} \otimes [\mathbf{S}_{ii}^{-1}]_{kl}$ and has size $m_im_j \times m_im_j$. Hence, $\mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1}$ is a matrix partitioned into blocks according to $m_im_j\mathbf{1}_K = [\underbrace{m_im_j, \ldots, m_im_j}_{K \text{ times}}]^\top$, both in rows and in columns. Matrices $\mathbf{S}_{ii}^{[k,l]}$ and $[\mathbf{S}_{ii}^{-1}]_{kl}$ are the (k,l)th blocks of $\mathbf{S}_{ii}$ and $\mathbf{S}_{ii}^{-1}$, respectively, and have size $m_i \times m_i$. The superscript notation $[k,l]$ is to remind that in the JISA context, $\mathbf{S}_{ii}^{[k,l]}$ is the covariance between sources $\mathbf{s}_i^{[k]}$ and $\mathbf{s}_i^{[l]}$ in data sets k and l, respectively. Since inverting $\mathbf{S}_{ii}$ mixes all data sets, we do not use this type of notation for sub-blocks of $\mathbf{S}_{ii}^{-1}$. Therefore, $\mathbf{S}_{ii}^{[k,l]} \triangleq [\mathbf{S}_{ii}]_{kl}$. In (2) we have introduced the commutation matrix $\boldsymbol{\mathcal{T}}_{P,Q} \in \mathbb{R}^{PQ\times PQ}$, where $\text{vec}\{\mathbf{M}^\dagger\} = \boldsymbol{\mathcal{T}}_{P,Q}\text{vec}\{\mathbf{M}\}$ for any $\mathbf{M} \in \mathbb{R}^{P\times Q}$ [28]. More properties of the commutation matrix can be found in Appendix A.

Matrix $\boldsymbol{\mathcal{H}}$ in (2) is always well-defined, since it is derived based on the assumption that $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ are invertible covariance matrices. This also implies that the identifiability results that are based on the analysis of $\boldsymbol{\mathcal{H}}$ are valid only for the case that $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ are invertible. For further discussion of what we define as admissible data, see Section 2.2 and Definition 2.3.

For the purpose of our analysis, we introduce a simplified notation, in which $\boldsymbol{\Theta} \triangleq \mathbf{S}_{jj}$ and $\boldsymbol{\Xi} \triangleq \mathbf{S}_{ii}$. Then, $\boldsymbol{\Theta}_{kl}$ and $[\boldsymbol{\Xi}^{-1}]_{kl}$ are $m_j \times m_j$ and $m_i \times m_i$ matrices, representing the (k,l)th blocks of $\boldsymbol{\Theta}$ and $\boldsymbol{\Xi}^{-1}$, according to the partitions $m_j\mathbf{1}_K = [\underbrace{m_j,\ldots,m_j}_{K \text{ times}}]^\top$ and $m_i\mathbf{1}_K = [\underbrace{m_i,\ldots,m_i}_{K \text{ times}}]^\top$, respectively. With this notation, matrix $\boldsymbol{\mathcal{H}}$ can now be written as

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \boldsymbol{\Theta} \boxplus \boldsymbol{\Xi}^{-1} & \mathbf{I}_K \otimes \boldsymbol{\mathcal{T}}_{m_j,m_i} \\ \mathbf{I}_K \otimes \boldsymbol{\mathcal{T}}_{m_i,m_j} & \boldsymbol{\Xi} \boxplus \boldsymbol{\Theta}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \boldsymbol{\mathcal{T}}_{m_i,m_j} \end{bmatrix} \underbrace{\begin{bmatrix} \boldsymbol{\Theta} \boxplus \boldsymbol{\Xi}^{-1} & \mathbf{I} \\ \mathbf{I} & \boldsymbol{\Theta}^{-1} \boxplus \boldsymbol{\Xi} \end{bmatrix}}_{\boldsymbol{\mathcal{H}}'} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \boldsymbol{\mathcal{T}}_{m_i,m_j}^\top \end{bmatrix} \tag{3}$$

where the factorization in the second step is due to Identity A.1 in Appendix A. Therefore, identifiability for this model consists in characterizing the sufficient and necessary conditions for the invertibility and thus positive-definiteness of

$$\boldsymbol{\mathcal{H}}' \triangleq \begin{bmatrix} \boldsymbol{\Theta} \boxplus \boldsymbol{\Xi}^{-1} & \mathbf{I} \\ \mathbf{I} & \boldsymbol{\Theta}^{-1} \boxplus \boldsymbol{\Xi} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} & \mathbf{I} \\ \mathbf{I} & \mathbf{S}_{jj}^{-1} \boxplus \mathbf{S}_{ii} \end{bmatrix}. \tag{4}$$

4.1 Analyzing $\boldsymbol{\mathcal{H}}$

For $\boldsymbol{\mathcal{H}}'$ to be positive-definite, we require that for any vector $\mathbf{x} \in \mathbb{K}^{2Km_im_j\times 1}$,

$$0 < \mathbf{x}^\top \boldsymbol{\mathcal{H}}'\mathbf{x} = \mathbf{x}^\top \mathbf{V}^\top \underbrace{\mathbf{V}\mathbf{x}}_{\mathbf{v}} = \mathbf{v}^\top \mathbf{v}\,. \tag{5}$$

Conversely, for $\boldsymbol{\mathcal{H}}'$ to be non-positive-definite, there must exist some non-zero $\mathbf{x} \in \mathbb{K}^{2Km_im_j\times 1}$ such that

$$0 = \mathbf{x}^\top \boldsymbol{\mathcal{H}}'\mathbf{x} = \mathbf{x}^\top \mathbf{V}^\top \underbrace{\mathbf{V}\mathbf{x}}_{\mathbf{v}\in\mathbb{K}^{2Km_im_j\times 1}} = \mathbf{v}^\top \mathbf{v} = \sum_{\alpha=1}^{2Km_im_j} |v_\alpha|^2 \quad \Leftrightarrow \quad v_\alpha = 0\;\forall\alpha \quad \Leftrightarrow \quad \mathbf{V}\mathbf{x} = \mathbf{0}\,. \tag{6}$$

4.1.1 Factorizing $\boldsymbol{\mathcal{H}}$

First, based on (5) and (6), we look for a meaningful factorization $\boldsymbol{\mathcal{H}}' = \mathbf{V}^\top\mathbf{V}$. We propose the following.

$$\boldsymbol{\mathcal{H}}' \overset{(4)}{=} \begin{bmatrix} \boldsymbol{\Theta} \boxplus \boldsymbol{\Xi}^{-1} & \mathbf{I} \\ \mathbf{I} & \boldsymbol{\Theta}^{-1} \boxplus \boldsymbol{\Xi} \end{bmatrix} \tag{7a}$$

$$\overset{(8)}{=} \begin{bmatrix} \boldsymbol{\Theta}^{\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxplus \boldsymbol{\Xi}^{-\frac{1}{2}\top}\boldsymbol{\Xi}^{-\frac{1}{2}} & \boldsymbol{\Theta}^{-\frac{1}{2}\top}\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxplus \boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Xi}^{-\frac{1}{2}} \\ \boldsymbol{\Theta}^{\frac{1}{2}}\boldsymbol{\Theta}^{-\frac{1}{2}} \boxplus \boldsymbol{\Xi}^{-\frac{1}{2}\top}\boldsymbol{\Xi}^{\frac{1}{2}\top} & \boldsymbol{\Theta}^{-\frac{1}{2}\top}\boldsymbol{\Theta}^{-\frac{1}{2}} \boxplus \boldsymbol{\Xi}^{\frac{1}{2}}\boldsymbol{\Xi}^{\frac{1}{2}\top} \end{bmatrix} \tag{7b}$$

$$\overset{\text{Identity A.5}}{=} \begin{bmatrix} (\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxbar \boldsymbol{\Xi}^{-\frac{1}{2}})^\top(\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxbar \boldsymbol{\Xi}^{-\frac{1}{2}}) & (\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxbar \boldsymbol{\Xi}^{-\frac{1}{2}})^\top(\boldsymbol{\Theta}^{-\frac{1}{2}} \boxbar \boldsymbol{\Xi}^{\frac{1}{2}\top}) \\ (\boldsymbol{\Theta}^{-\frac{1}{2}} \boxbar \boldsymbol{\Xi}^{\frac{1}{2}\top})^\top(\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxbar \boldsymbol{\Xi}^{-\frac{1}{2}}) & (\boldsymbol{\Theta}^{-\frac{1}{2}} \boxbar \boldsymbol{\Xi}^{\frac{1}{2}\top})^\top(\boldsymbol{\Theta}^{-\frac{1}{2}} \boxbar \boldsymbol{\Xi}^{\frac{1}{2}\top}) \end{bmatrix} \tag{7c}$$

$$= \underbrace{\begin{bmatrix} (\boldsymbol{\Theta}^{\frac{1}{2}\top} \boxbar \boldsymbol{\Xi}^{-\frac{1}{2}})^\top \\ (\boldsymbol{\Theta}^{-\frac{1}{2}} \boxbar \boldsymbol{\Xi}^{\frac{1}{2}\top})^\top \end{bmatrix}}_{\mathbf{V}^\top} \underbrace{\begin{bmatrix} \boldsymbol{\Theta}^{\frac{1}{2}\top} \boxbar \boldsymbol{\Xi}^{-\frac{1}{2}} & \boldsymbol{\Theta}^{-\frac{1}{2}} \boxbar \boldsymbol{\Xi}^{\frac{1}{2}\top} \end{bmatrix}}_{\mathbf{V}_{Km_im_j\times 2Km_im_j}} = \mathbf{V}^\top\mathbf{V} \tag{7d}$$

The first equality repeats the definition of $\boldsymbol{\mathcal{H}}'$ in (4). The second equality uses the square root factorization of a symmetric matrix, which we define as

$$\mathbf{S}_{jj} = \mathbf{S}_{jj}^{\frac{1}{2}}\mathbf{S}_{jj}^{\frac{1}{2}\top} = \boldsymbol{\Theta}^{\frac{1}{2}}\boldsymbol{\Theta}^{\frac{1}{2}\top} = \boldsymbol{\Theta} \quad \Leftrightarrow \quad \mathbf{S}_{jj}^{-1} = \mathbf{S}_{jj}^{-\frac{1}{2}\top}\mathbf{S}_{jj}^{-\frac{1}{2}} = \boldsymbol{\Theta}^{-\frac{1}{2}\top}\boldsymbol{\Theta}^{-\frac{1}{2}} = \boldsymbol{\Theta}^{-1} \tag{8}$$

The third equality follows from Identity A.5 in Appendix A, which leads directly to the desired factorization in the fourth step.

4.1.2 Find $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{V}\mathbf{x} = \mathbf{0}$

Next, we find a non-zero vector $\mathbf{x} \in \mathbb{K}^{2Km_im_j \times 1}$ such that $\mathbf{V}\mathbf{x} = \mathbf{0}$. W.l.o.g., we look for $\mathbf{x}$ in the general form

$$\mathbf{x} = \begin{bmatrix} \mathrm{vec}\{\mathbf{M}^{[1]}\} \\ \vdots \\ \mathrm{vec}\{\mathbf{M}^{[K]}\} \\ -\mathrm{vec}\{\mathbf{N}^{[1]}\} \\ \vdots \\ -\mathrm{vec}\{\mathbf{N}^{[K]}\} \end{bmatrix} = \begin{bmatrix} \mathrm{vecbd}_\varsigma\{\mathbf{M}\} \\ -\mathrm{vecbd}_\varsigma\{\mathbf{N}\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ -\boldsymbol{\nu} \end{bmatrix} \tag{9}$$

where

$$\boldsymbol{\mu} = \mathrm{vecbd}_\varsigma\{\mathbf{M}\} = \begin{bmatrix} \mathrm{vec}\{\mathbf{M}^{[1]}\} \\ \vdots \\ \mathrm{vec}\{\mathbf{M}^{[K]}\} \end{bmatrix}, \, \mathbf{M} \triangleq \bigoplus_{k=1}^{K} \mathbf{M}^{[k]} = \mathbf{M}\,, \, \mathbf{M}^{[k]} \in \mathbb{K}^{m_i \times m_j}\,,$$
$$\boldsymbol{\nu} = \mathrm{vecbd}_\varsigma\{\mathbf{N}\} = \begin{bmatrix} \mathrm{vec}\{\mathbf{N}^{[1]}\} \\ \vdots \\ \mathrm{vec}\{\mathbf{N}^{[K]}\} \end{bmatrix}, \, \mathbf{N} \triangleq \bigoplus_{k=1}^{K} \mathbf{N}^{[k]} = \mathbf{N}\,, \, \mathbf{N}^{[k]} \in \mathbb{K}^{m_i \times m_j}$$

The "vecbd" operator is defined in Definition A.2 in Appendix A. The symbol ς stands for the block-partition $m_i\mathbf{1}_K \times m_j\mathbf{1}_K$. Setting $\mathbf{V}\mathbf{x} = 0$ implies that

$$\left[\mathbf{S}_{jj}^{\frac{1}{2}\top} \boxminus \mathbf{S}_{ii}^{-\frac{1}{2}} \quad \mathbf{S}_{jj}^{-\frac{1}{2}} \boxminus \mathbf{S}_{ii}^{\frac{1}{2}\top} \right] \begin{bmatrix} \mathrm{vecbd}_\varsigma\{\mathbf{M}\} \\ -\mathrm{vecbd}_\varsigma\{\mathbf{N}\} \end{bmatrix} = \mathbf{0} \tag{10}$$

for some non-zero $\mathrm{vecbd}_\varsigma\{\mathbf{M}\}$ and $\mathrm{vecbd}_\varsigma\{\mathbf{N}\}$. Equality (10) can be rewritten as

$$(\mathbf{S}_{jj}^{\frac{1}{2}\top} \boxminus \mathbf{S}_{ii}^{-\frac{1}{2}})\mathrm{vecbd}_\varsigma\{\mathbf{M}\} = (\mathbf{S}_{jj}^{-\frac{1}{2}} \boxminus \mathbf{S}_{ii}^{\frac{1}{2}\top})\mathrm{vecbd}_\varsigma\{\mathbf{N}\} \tag{11}$$

We now turn to finding these $\mathrm{vecbd}_\varsigma\{\mathbf{M}\}$ and $\mathrm{vecbd}_\varsigma\{\mathbf{N}\}$ (alternatively: $\{\mathbf{M}^{[k]}\}_{k=1}^K$ and $\{\mathbf{N}^{[k]}\}_{k=1}^K$). Using Identity A.3 in Appendix A, equality (11) can be rewritten as

$$\mathrm{vec}\{\mathbf{S}_{ii}^{-\frac{1}{2}}\mathbf{M}\mathbf{S}_{jj}^{\frac{1}{2}}\} = \mathrm{vec}\{\mathbf{S}_{ii}^{\frac{1}{2}\top}\mathbf{N}\mathbf{S}_{jj}^{-\frac{1}{2}\top}\}\,. \tag{12}$$

Removing the "vec" notation,

$$\mathbf{S}_{ii}^{-\frac{1}{2}}\mathbf{M}\mathbf{S}_{jj}^{\frac{1}{2}} = \mathbf{S}_{ii}^{\frac{1}{2}\top}\mathbf{N}\mathbf{S}_{jj}^{-\frac{1}{2}\top}\,. \tag{13}$$

Since $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ are invertible, the latter is equivalent to

$$\boxed{\begin{aligned} &\mathbf{M}^{[k]}\mathbf{S}_{jj}^{[k,l]} = \mathbf{S}_{ii}^{[k,l]}\mathbf{N}^{[l]} \quad \in \mathbb{K}^{m_i \times m_j} \quad \forall k,l && \text{(14a)} \\ &\mathbf{M}\mathbf{S}_{jj} = \mathbf{S}_{ii}\mathbf{N} \quad , \, \mathbf{M} \triangleq \bigoplus_{k=1}^{K} \mathbf{M}^{[k]} \quad , \, \mathbf{N} \triangleq \bigoplus_{k=1}^{K} \mathbf{N}^{[k]} && \text{(14b)} \end{aligned}}$$

Hence, our goal is to find non-zero $\{\mathbf{M}^{[k]}\}_{k=1}^K$ and/or $\{\mathbf{N}^{[k]}\}_{k=1}^K$ for which equality (14) holds.

4.1.3 From $\mathbf{M}\mathbf{S}_{jj} = \mathbf{S}_{ii}\mathbf{N}$ to $\mathbf{L}\mathbf{R} = \mathbf{P}\mathbf{L}$

The identifiability problem (14) can further be simplified into characterizing the non-trivial solutions to

$$\boxed{\begin{aligned} &\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{L}^{[l]} \quad \in \mathbb{K}^{P \times R} \quad \forall k,l && \text{(15a)} \\ &\mathbf{L}\mathbf{R} = \mathbf{P}\mathbf{L} \quad , \, \mathbf{L} \triangleq \bigoplus_{k=1}^{K} \mathbf{L}^{[k]} && \text{(15b)} \end{aligned}}$$

where $\mathbf{P} \in \mathbb{K}^{KP \times KP}$ and $\mathbf{R} \in \mathbb{K}^{KR \times KR}$ are normalized versions of $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ such that their (k,k)th main-diagonal blocks are equal to the identity:

$$\mathbf{P}^{[k,k]} = \mathbf{I}_P \quad \text{and} \quad \mathbf{R}^{[k,k]} = \mathbf{I}_R \tag{16}$$

and $P = m_i$, $R = m_j$. The normalization scheme that leads to (16) is explained in Appendix D. The derivation of (15) is explained in Appendix E. The positive definite matrices $\mathbf{P}$ and $\mathbf{R}$ are partitioned similarly to $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ into $K \times K$ blocks such that $\mathbf{P}^{[k,l]} \in \mathbb{K}^{P\times P}$ and $\mathbf{R}^{[k,l]} \in \mathbb{K}^{R\times R}$. Accordingly, $\mathbf{L}^{[k]} \in \mathbb{K}^{P\times R}$, $\mathbf{L} \triangleq \operatorname{bdiag}\{\mathbf{L}^{[1]}, \ldots, \mathbf{L}^{[K]}\} = \bigoplus_{k=1}^{K} \mathbf{L}^{[k]}$. Problem (15) is simpler than (14) since $\mathbf{L}$ replaces both $\mathbf{M}$ and $\mathbf{N}$, thus cutting by half the number of unknowns. The problem can now be reformulated as finding the minimal conditions on $\mathbf{P}$ and $\mathbf{R}$ such that $\{\mathbf{L}^{[k]}\}_{k=1}^{K}$ are not all zero and (15) holds, while respecting the admissibility constraints in Definition 2.3.

4.2 Main Result: JISA Identifiability

We now apply Lemma 3–4 to (15) and then undo the normalization. For $\mathbf{R}^{[k,l]}$, $\mathbf{P}^{[k,l]}$ with structure as in (68) and $m_i = m_j$, it follows from Lemma 3–4 that $\mathbf{L}^{[k]} = \lambda \mathbf{O}^{[k]}\ \forall k \in \mathcal{D}$. Hence,

$$\begin{aligned}\lambda \mathbf{O}^{[k]} \cdot \boldsymbol{\Omega}_{jj}^{[k]} \mathbf{S}_{jj}^{[k,l]} \boldsymbol{\Omega}_{jj}^{[l]\top} &= \boldsymbol{\Omega}_{ii}^{[k]} \mathbf{S}_{ii}^{[k,l]} \boldsymbol{\Omega}_{ii}^{[l]\top} \cdot \lambda \mathbf{O}^{[l]} \\ \Rightarrow \quad \mathbf{S}_{jj}^{[k,l]} &= \underbrace{\boldsymbol{\Omega}_{jj}^{-[k]} \mathbf{O}^{[k]\top} \boldsymbol{\Omega}_{ii}^{[k]}}_{\boldsymbol{\Psi}^{[k]}} \mathbf{S}_{ii}^{[k,l]} \underbrace{\boldsymbol{\Omega}_{ii}^{[l]\top} \mathbf{O}^{[l]} \boldsymbol{\Omega}_{jj}^{-[l]\top}}_{\boldsymbol{\Psi}^{[l]\top}}\end{aligned}$$

This link between matrices can be regarded as a generalized similarity or equivalence transformation, as defined in Definition 2.2; see also Vía et al. [11, Definition 1]. For two datasets with $\mathbf{R}^{[k,l]}$, $\mathbf{P}^{[k,l]}$ with structure as in (69), Lemma 4 states that there always exists a non-trivial solution to (15). Hence, data with structure (69) are always non-identifiable. In all other cases, the model is identifiable.

Accordingly, we identify two types of scenarios associated with non-identifiability of the basic JISA model.

Scenario 4.1. The first type of non-identifiability is associated with a pair of sources whose covariance matrices have the structure

$$\mathbf{S}_{ii} = \left[\begin{array}{c|c} \mathbf{S}_{ii}^{[1:D,1:D]} & \mathbf{0}_{Dm_i\times(K-D)m_i} \\ \hline \mathbf{0}_{(K-D)m_i\times Dm_i} & \mathbf{S}_{ii}^{[D+1:K,D+1:K]} \end{array}\right] \quad \text{and} \quad \mathbf{S}_{jj} = \left[\begin{array}{c|c} \mathbf{S}_{jj}^{[1:D,1]} & \mathbf{0}_{Dm_j\times(K-D)m_j} \\ \hline \mathbf{0}_{(K-D)m_j\times Dm_j} & \mathbf{S}_{jj}^{[D+1:K,D+1:K]} \end{array}\right] \tag{17}$$

In this case, the model is non-identifiable iff $m_i = m_j$ and the sources are linked by

$$\mathbf{S}_{jj}^{[1:D,1:D]} = \operatorname{bdiag}\{\boldsymbol{\Psi}^{[1]}, \ldots, \boldsymbol{\Psi}^{[D]}\} \mathbf{S}_{ii}^{[1:D,1:D]} \operatorname{bdiag}^{\top}\{\boldsymbol{\Psi}^{[1]}, \ldots, \boldsymbol{\Psi}^{[D]}\} \tag{18}$$

for any D invertible $m_i \times m_i$ matrices $\boldsymbol{\Psi}^{[k]}$, $k = 1, \ldots, D$. Equation (18) is a generalized similarity transformation, hence an equivalence relation, in the sense of Definition 2.2, between $\mathbf{S}_{ii}^{[1:D,1:D]}$ and $\mathbf{S}_{jj}^{[1:D,1:D]}$.

Scenario 4.2. The second type of non-identifiability is associated with a pair of sources whose covariance matrices are of possibly different size, $m_i \neq m_j$, and structure

$$\mathbf{S}_{ii} = \left[\begin{array}{ccc|c} \mathbf{S}_{ii}^{[1,1]} & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0}_{Dm_i\times(K-D)m_i} \\ \mathbf{0} & \cdots & \mathbf{S}_{ii}^{[D,D]} & \\ \hline & \mathbf{0}_{(K-D)m_i\times Dm_i} & & \mathbf{S}_{ii}^{[D+1:K,D+1:K]} \end{array}\right] \tag{19a}$$

$$\mathbf{S}_{jj} = \left[\begin{array}{ccc|c} \mathbf{S}_{jj}^{[1,1]} & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0}_{Dm_j\times(K-D)m_j} \\ \mathbf{0} & \cdots & \mathbf{S}_{jj}^{[D,D]} & \\ \hline & \mathbf{0}_{(K-D)m_j\times Dm_j} & & \mathbf{S}_{jj}^{[D+1:K,D+1:K]} \end{array}\right] \tag{19b}$$

In this case, the model is always non-identifiable.

Theorem 4.1 (JISA non-identifiability)**.** *The JISA model is not identifiable iff there exists at least one pair (i, j) of positive definite covariance matrices $\mathbf{S}_{jj}$ and $\mathbf{S}_{ii}$, admissible by Definition 2.3, for which either*

1. *structure (19) holds for $1 \leq D \leq K$,*

or

2. *structure (17) holds for $2 \leq D \leq K$, $m_i^{[k]} = m_j^{[k]}\ \forall k \in \mathcal{D}$, and*

$$\mathbf{S}_{jj}^{[1:D,1:D]} = \operatorname{bdiag}\{\boldsymbol{\Psi}^{[1]}, \ldots, \boldsymbol{\Psi}^{[D]}\} \mathbf{S}_{ii}^{[1:D,1:D]} \operatorname{bdiag}^{\top}\{\boldsymbol{\Psi}^{[1]}, \ldots, \boldsymbol{\Psi}^{[D]}\}$$

where $\boldsymbol{\Psi}^{[k]}$ are arbitrary invertible $m_i^{[k]} \times m_i^{[k]}$ matrices.

5 Discussion

In this paper, we presented new results in algebra and in signal processing. This was achieved by analysing the FIM of a recently-proposed source separation model that is inspired by data fusion. Therefore, one of our messages in this work is that by formulating new ways in which data sets can interact, we obtain new types of algebraic structures, and the theoretical analysis of these algebraic structures yields new insights and contributions that go beyond their community of origin.

This algebraic result, formulated in several lemmas, can be regarded as a generalization to and a variation of Schur's lemma on irreducible representations. As an application, we have used this lemma to derive the non-generic uniqueness and identifiability conditions of JISA, when SOS are involved and the mixing matrices are all invertible. This model, as well as the corresponding lemmas, can be extended by further relaxing some of the numerical and structural assumptions.

From a data fusion perspective, the JISA model is non-identifiable in two main scenarios: first, if there exists at least one pair of sources with equivalent subspaces, in the generalized sense. Second, if there exists a pair of sources with no counterparts in the other data sets. All other scenarios are identifiable, if the data are admissible (as long as the basic JISA model assumptions hold, of course). This implies that JISA can be used for data fusion even if there are only very few links among corresponding sources in different data sets.

The notion of "generalized similarity" may be used as a multiset extension of the concept of "Kruskal rank" [29] to datasets that cannot be stacked in a single array (tensor), yet having a multi-way nature. Our results are related to the concept of k'-rank, introduced in [30, Definition 2.3]. The results in this paper allow to further generalize these concepts. A potential impact is using this generalization to derive the uniqueness of more elaborate coupled and multi-way models.

Finally, Table 1 provides further insights into our new results by comparing the original and new lemmas. "Variation" in the first row implies that we use symmetric matrices instead of irreducible representations of symmetry groups, as in the original formulation by Schur. Table 1 clarifies why we call the new formulation "generalized": when all commuting matrices are forced to be identical, we obtain the original Schur formulation.

		Multiset analogue (new)	Schur's lemma (variation)
First lemma	Input data	$k, l = 1, \dots, K$ $\mathbf{M}^{[k]}$, $\mathbf{C}^{[k,l]}$ $P \times P$ $\mathbf{C}^{[k,l]} = (\mathbf{C}^{[l,k]})^\top$	$q = 1, \dots, Q$ $\mathbf{M}$, $\mathbf{C}^{(q)}$ $P \times P$ $\mathbf{C}^{(q)} = (\mathbf{C}^{(q)})^\top$
	Commutation Non-uniqueness Comm. if $\mathbf{M}^{[k]} = \mathbf{M}\ \forall k$	$\mathbf{M}^{[k]}\mathbf{C}^{[k,l]} = \mathbf{C}^{[k,l]}\mathbf{M}^{[l]}\ \forall k, l$ $\mathbf{M}^{[k]} = \mu \mathbf{I}_P\ \forall k,\ \mu \geq 0$ $\mathbf{M}\mathbf{C}^{[k,l]} = \mathbf{C}^{[k,l]}\mathbf{M}\ \forall k, l$	$\mathbf{M}\mathbf{C}^{(q)} = \mathbf{C}^{(q)}\mathbf{M}\ \forall q$ $\mathbf{M} = \mu \mathbf{I}_P,\ \mu \geq 0$
Second lemma	Input data	$\mathbf{M}^{[k]}$ $P' \times P$ $\mathbf{C}^{[k,l]}$ $P \times P$, $\mathbf{C}'^{[k,l]}$ $P' \times P'$ $P \times P$, $\mathbf{C}'^{(q)}$ $P' \times P'$	$\mathbf{M}$ $P' \times P$ $\mathbf{C}^{(q)}$
	Commutation Non-uniqueness Comm. if $\mathbf{M}^{[k]} = \mathbf{M}\ \forall k$	$\mathbf{M}^{[k]}\mathbf{C}^{[k,l]} = \mathbf{C}'^{[k,l]}\mathbf{M}^{[l]}\ \forall k, l$ $\mathbf{M}^{[k]} = \mathbf{0}_{P' \times P}$ or $\mathbf{M}^{[k]} = \lambda \mathbf{O}^{[k]}\ \forall k$ $\mathbf{M}\mathbf{C}^{[k,l]} = \mathbf{C}'^{[k,l]}\mathbf{M}\ \forall k, l$	$\mathbf{M}\mathbf{C}^{(q)} = \mathbf{C}'^{(q)}\mathbf{M}\ \forall q$ $\mathbf{M} = \mathbf{0}_{P' \times P}$ or $\mathbf{M} = \lambda \mathbf{O}$

Table 1: Comparison of Schur's lemma and its proposed multiset analogue, strict sense

A Some Algebraic Properties

For ease of reference, we list some useful algebraic properties. Properties that are not proved below can be found in [31, 32, 28]. A glossary of notations is given in Table 2.

Product Name	Notation	LaTex Command
Hadamard	$\circledast$	`\had`
Khatri-Rao columnwise	$\odot$	`\khat`
Khatri-Rao block-columnwise	$\boxbar$	`\khatcb`
Khatri-Rao for partitioned matrices / block-Kronecker	$\boxplus$	`\khatb`

Table 2: Glossary

For any matrices $\mathbf{M}, \mathbf{N}, \mathbf{P}, \mathbf{Q}$ (with appropriate dimensions),

$$(\mathbf{N} \otimes \mathbf{M})(\mathbf{P} \otimes \mathbf{Q}) = \mathbf{NP} \otimes \mathbf{MQ} \tag{20a}$$

$$(\mathbf{N} \otimes \mathbf{M})^\top = \mathbf{N}^\top \otimes \mathbf{M}^\top \tag{20b}$$

$$\mathrm{vec}\{\mathbf{MQN}\} = (\mathbf{N}^\top \otimes \mathbf{M})\mathrm{vec}\{\mathbf{Q}\} \tag{20c}$$

$$\mathrm{tr}\{\mathbf{PQ}\} = \mathrm{tr}\{\mathbf{QP}\} \tag{20d}$$

$$\mathrm{tr}\{\mathbf{P}^\top\mathbf{Q}\} = \mathrm{vec}^\dagger\{\mathbf{P}\}\mathrm{vec}\{\mathbf{Q}\} \tag{20e}$$

$$\det(\mathbf{MN}) = \det(\mathbf{NM})\,. \tag{20f}$$

For any two matrices $\mathbf{M}_{M\times P}$ and $\mathbf{N}_{N\times Q}$,

$$\boldsymbol{\mathcal{T}}_{M,N}(\mathbf{N}\otimes\mathbf{M}) = (\mathbf{M}\otimes\mathbf{N})\boldsymbol{\mathcal{T}}_{P,Q}\,. \tag{21a}$$

Identity A.1.

$$(\mathbf{I}\otimes\boldsymbol{\mathcal{T}}_{m_i,m_j})(\mathbf{A}\boxplus\mathbf{B})(\mathbf{I}\otimes\boldsymbol{\mathcal{T}}^\top_{m_i,m_j}) = \mathbf{B}\boxplus\mathbf{A} \tag{22}$$

Proof of Identity A.1.

$$\begin{aligned}
&(\mathbf{I}\otimes\boldsymbol{\mathcal{T}}_{m_i,m_j})(\mathbf{A}\boxplus\mathbf{B})(\mathbf{I}\otimes\boldsymbol{\mathcal{T}}^\top_{m_i,m_j}) \\
&= \begin{bmatrix} \boldsymbol{\mathcal{T}}_{m_i,m_j} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\mathcal{T}}_{m_i,m_j} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11}\otimes\mathbf{B}_{11} & \cdots & \mathbf{A}_{1K}\otimes\mathbf{B}_{1K} \\ \vdots & & \vdots \\ \mathbf{A}_{K1}\otimes\mathbf{B}_{K1} & \cdots & \mathbf{A}_{KK}\otimes\mathbf{B}_{KK} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{T}}^\top_{m_i,m_j} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\mathcal{T}}^\top_{m_i,m_j} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\mathcal{T}}_{m_i,m_j}(\mathbf{A}_{11}\otimes\mathbf{B}_{11})\boldsymbol{\mathcal{T}}^\top_{m_i,m_j} & \cdots & \boldsymbol{\mathcal{T}}_{m_i,m_j}(\mathbf{A}_{1K}\otimes\mathbf{B}_{1K})\boldsymbol{\mathcal{T}}^\top_{m_i,m_j} \\ \vdots & & \vdots \\ \boldsymbol{\mathcal{T}}_{m_i,m_j}(\mathbf{A}_{K1}\otimes\mathbf{B}_{K1})\boldsymbol{\mathcal{T}}^\top_{m_i,m_j} & \cdots & \boldsymbol{\mathcal{T}}_{m_i,m_j}(\mathbf{A}_{KK}\otimes\mathbf{B}_{KK})\boldsymbol{\mathcal{T}}^\top_{m_i,m_j} \end{bmatrix} \\
&\overset{(21a)}{=} \begin{bmatrix} \mathbf{B}_{11}\otimes\mathbf{A}_{11} & \cdots & \mathbf{B}_{1K}\otimes\mathbf{A}_{1K} \\ \vdots & & \vdots \\ \mathbf{B}_{K1}\otimes\mathbf{A}_{K1} & \cdots & \mathbf{B}_{KK}\otimes\mathbf{A}_{KK} \end{bmatrix} = \mathbf{B}\boxplus\mathbf{A}
\end{aligned} \tag{23}$$

□

A.1 diag, bdiag, vecd, vecbd operators

In order to avoid confusion with the vecd and vecbd operators, we define

$$\mathrm{diag}\{\mathbf{X}\} \triangleq \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_{KK} \end{bmatrix} = \texttt{diag(diag(X))}\ ,\ \mathbf{X}\in\mathbb{R}^{K\times K}$$

$$\mathrm{diag}\{\mathbf{x}\} \triangleq \begin{bmatrix} x_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_K \end{bmatrix} = \texttt{diag(x)}\ ,\ \mathbf{x}\in\mathbb{R}^{K\times 1}$$

Now, let $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^\top$, $\sum_{k=1}^K \alpha_k = \alpha$ and similarly for β. Then,

$$\mathrm{bdiag}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\} \triangleq \begin{bmatrix} \mathbf{X}_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{KK} \end{bmatrix} = \bigoplus_{k=1}^{K} \mathbf{X}_{kk} = \mathrm{bdiag}\{\mathbf{X}_{11}, \ldots, \mathbf{X}_{KK}\} \quad , \mathbf{X}_{kk} \in \mathbb{R}^{\alpha_k \times \beta_k}$$
$$\mathrm{bdiag}_{\boldsymbol{\alpha}}\{\mathbf{X}\} \triangleq \mathrm{bdiag}_{\boldsymbol{\alpha}\times\boldsymbol{\alpha}}\{\mathbf{X}\} \ , \ \mathbf{X} \in \mathbb{R}^{\alpha\times\alpha}$$

Definition A.1 (vecd Operator)**.** For any square matrix $\mathbf{X} \in \mathbb{R}^{K\times K}$ with entries x_{ij}, $i, j \in 1, \ldots, K$, define the operator

$$\mathrm{vecd}\{\mathbf{X}\} \triangleq \begin{bmatrix} x_{11} \\ \vdots \\ x_{KK} \end{bmatrix} = \mathtt{diag(X)} \ \in \mathbb{R}^{K\times 1} . \tag{24}$$

That is, $\mathrm{vecd}\{\mathbf{X}\}$ is a vector that consists only of the entries on the diagonal of $\mathbf{X}$.

Definition A.2 (vecbd Operator)**.** For any rectangular matrix $\mathbf{X} \in \mathbb{R}^{\alpha\times\beta}$ partitioned into K rows and K columns such that its (i, j)th block is $\mathbf{X}_{ij} \in \mathbb{R}^{\alpha_k\times\beta_k}$, $i, j \in 1, \ldots, K$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^\top$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^\top$, $\alpha = \sum_{k=1}^K \alpha_k$, $\beta = \sum_{k=1}^K \beta_k$, define the operator

$$\mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\} \triangleq \begin{bmatrix} \mathrm{vec}\{\mathbf{X}_{11}\} \\ \vdots \\ \mathrm{vec}\{\mathbf{X}_{KK}\} \end{bmatrix} \in \mathbb{R}^{(\sum_{k=1}^K \alpha_k\beta_k)\times 1} \quad \neq \mathrm{vec}\{\mathrm{bdiag}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\}\} \tag{25}$$

That is, $\mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\}$ is a vector that consists only of the (vectorized) entries of the block-diagonal of $\mathbf{X}$, where the rows of $\mathbf{X}$ are partitioned according to $\boldsymbol{\alpha}$ and the columns by $\boldsymbol{\beta}$. If $\boldsymbol{\alpha} = \boldsymbol{\beta}$ then we can write

$$\mathrm{vecbd}_{\boldsymbol{\alpha}}\{\mathbf{X}\} \triangleq \mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\alpha}}\{\mathbf{X}\} . \tag{26}$$

A.2 Khatri-Rao, Kronecker, Tensor Matricization and Vectorization

Tensors can be written as multidimensional arrays, matricized or vectorized. Even within these representations, there are variations. Consider a third-order tensor whose representation in multilinear products is

$$\mathcal{T} = \mathcal{D} \times_1 \mathbf{A}' \times_2 \mathbf{B}' \times_3 \mathbf{C}' \quad \text{(CPD)} \tag{27a}$$
$$\mathcal{T} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \quad \text{(Tucker)} \tag{27b}$$

where $\mathcal{D}$ is a tensor with diagonal core $\mathbf{d}$ [33, Table I, Eq. (4)]. Then, its vector representation can be written as [33, Table III]

$$\mathrm{vec}\{\mathcal{T}\} = (\mathbf{C}' \odot \mathbf{B}' \odot \mathbf{A}')\mathbf{d} \quad \text{(CPD)} \tag{28a}$$
$$\mathrm{vec}\{\mathcal{T}\} = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\mathrm{vec}\{\mathcal{G}\} \quad \text{(Tucker)} \tag{28b}$$

If now $\mathcal{T}$ is a second-order tensor, then (27) reduces to

$$\mathbf{T} = \mathbf{D} \times_1 \mathbf{A}' \times_2 \mathbf{B}' = \mathbf{A}'\mathbf{D}\mathbf{B}'^\top \ , \ \mathbf{D} = \mathrm{diag}\{\mathbf{d}\} \tag{29a}$$
$$\mathbf{T} = \mathbf{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \ = \mathbf{A}\mathbf{G}\mathbf{B}^\top \tag{29b}$$

and (28) to

$$\mathrm{vec}\{\mathbf{T}\} = (\mathbf{B}' \odot \mathbf{A}')\mathbf{d} \tag{30a}$$
$$\mathrm{vec}\{\mathbf{T}\} = (\mathbf{B} \otimes \mathbf{A})\mathrm{vec}\{\mathbf{G}\} \tag{30b}$$

Combining (29) with (30), and using the notation $\mathbf{d} = \mathrm{vecd}\{\mathbf{D}\}$ from Definition A.1, we obtain

$$\mathrm{vec}\{\mathbf{T}\} = \overbrace{(\mathbf{B}' \odot \mathbf{A}')\mathrm{vecd}\{\mathbf{D}\} = \mathrm{vec}\{\mathbf{A}'\mathbf{D}\mathbf{B}'^\top\}}^{\text{Only for } \mathbf{D}\in\mathrm{diag}} \tag{31a}$$
$$\mathrm{vec}\{\mathbf{T}\} = \underbrace{(\mathbf{B} \otimes \mathbf{A})\mathrm{vec}\{\mathbf{G}\} = \mathrm{vec}\{\mathbf{A}\mathbf{G}\mathbf{B}^\top\}}_{\text{Well-known identity } \forall\mathbf{G}} \tag{31b}$$

The equalities in (31) are summarized in the following identity

Identity A.2. For any $\mathbf{X} \in \text{diag}$, and matrices $\mathbf{A}$ and $\mathbf{B}$ with appropriate dimensions,

$$(\mathbf{B} \odot \mathbf{A})\text{vecd}\{\mathbf{X}\} \overset{\mathbf{X}\in\text{diag}}{=} (\mathbf{B} \otimes \mathbf{A})\text{vec}\{\mathbf{X}\} \overset{\forall\mathbf{X}}{=} \text{vec}\{\mathbf{A}\mathbf{X}\mathbf{B}^\top\} \tag{32}$$

where the operator "vecd$\{\cdot\}$" was defined in Definition A.1. The second equality is true for any $\mathbf{X}$, and the first one only for $\mathbf{X}$ diagonal.

Identity A.2 appears in Brewer [34, Table III, T3.13], as well as in Liu and Trenkler [35, Equation (27)].

Similarly to the argumentations in (27)–(31), we can also vectorize a Tucker format of a second-order tensor when the core tensor is block-diagonal.

Identity A.3. Let $\mathbf{A} \in \mathbb{R}^{\mu\times\alpha}$ and $\mathbf{B} \in \mathbb{R}^{\nu\times\beta}$ be two matrices partitioned into K column blocks of dimensions $\mu \times \alpha_k$ and $\nu \times \beta_k$, respectively, $\alpha = \sum_{k=1}^K \alpha_k$, $\beta = \sum_{k=1}^K \beta_k$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^\top$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^\top$, as follows,

$$\begin{aligned}\mathbf{A} &= \left[\ \mathbf{A}_1 \mid \cdots \mid \mathbf{A}_K\ \right], \quad \mathbf{A}_k \in \mathbb{R}^{\mu\times\alpha_k}\\ \mathbf{B} &= \left[\ \mathbf{B}_1 \mid \cdots \mid \mathbf{B}_K\ \right], \quad \mathbf{B}_k \in \mathbb{R}^{\nu\times\beta_k}\end{aligned} \tag{33}$$

and $\mathbf{X} = \bigoplus_{k=1}^K \mathbf{X}_{kk} \in \mathbb{R}^{\alpha\times\beta}$, $\mathbf{X}_{kk} \in \mathbb{R}^{\alpha_k\times\beta_k}$. Then

$$(\mathbf{B} \boxdot \mathbf{A})\text{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\} = \text{vec}\{\mathbf{A}\mathbf{X}\mathbf{B}^\top\}, \tag{34}$$

where the operator "vecbd$_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\cdot\}$" was defined in Definition A.2 and "$\boxdot$" in Table 2.

Remark 3. Identity A.3 is a generalization of Identity A.2.

Proof of Identity A.3 and Equation (34). Apart from the Tucker format vectorization, which is a constructive proof, we can also prove directly. The following proof is based on the fact that

$$\mathbf{A}_k = \mathbf{A}\mathbf{J}_{\alpha_k} \tag{35}$$

where $\mathbf{J}_{\alpha_k} = \begin{bmatrix} \mathbf{0} \\ \hline \mathbf{I}_{\alpha_k} \\ \hline \mathbf{0} \end{bmatrix} \in \mathbb{R}^{\alpha\times\alpha_k}$ is a matrix of zeros, with $\mathbf{I}_{\alpha_k}$ at the rows required to extract the columns pertaining to $\mathbf{A}_k$ from $\mathbf{A}$. We define a similar matrix for $\mathbf{B}$. Then,

$$\begin{aligned}(\mathbf{B} \boxdot \mathbf{A})\text{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\} &= \left[\ \mathbf{B}\mathbf{J}_{\beta_1} \otimes \mathbf{A}\mathbf{J}_{\alpha_1} \mid \cdots \mid \mathbf{B}\mathbf{J}_{\beta_K} \otimes \mathbf{A}\mathbf{J}_{\alpha_K}\ \right] \begin{bmatrix} \text{vec}\{\mathbf{X}_{11}\} \\ \vdots \\ \text{vec}\{\mathbf{X}_{KK}\} \end{bmatrix}\\ &= \sum_{k=1}^K (\mathbf{B}^\top\mathbf{J}_{\beta_k} \otimes \mathbf{A}\mathbf{J}_{\alpha_k})\text{vec}\{\mathbf{X}_{kk}\} \overset{(20c)}{=} \sum_{k=1}^K \text{vec}\{\mathbf{A}\mathbf{J}_{\alpha_k}\mathbf{X}_{kk}\mathbf{J}_{\beta_k}^\top\mathbf{B}\}\\ &= \text{vec}\{\mathbf{A}\underbrace{\Big(\sum_{k=1}^K \mathbf{J}_{\alpha_k}\mathbf{X}_{kk}\mathbf{J}_{\beta_k}^\top\Big)}_{\bigoplus_{k=1}^K \mathbf{X}_{kk} = \mathbf{X}}\mathbf{B}^\top\} = \text{vec}\{\mathbf{A}\mathbf{X}\mathbf{B}^\top\}\end{aligned} \tag{36}$$

$\square$

Identity A.4. Consider four matrices $\mathbf{A} \in \mathbb{R}^{\mu\times\xi}$, $\mathbf{B} \in \mathbb{R}^{\nu\times\zeta}$, $\mathbf{C} \in \mathbb{R}^{\alpha\times\eta}$, $\mathbf{D} \in \mathbb{R}^{\gamma\times\delta}$. Then,

$$\underbrace{\overbrace{\underbrace{\mathbf{A}^\top\mathbf{C}}_{\Rightarrow\mu=\alpha}}^{\xi\times\eta} \circledast \overbrace{\underbrace{\mathbf{B}^\top\mathbf{D}}_{\Rightarrow\nu=\gamma}}^{\zeta\times\delta}}_{\Rightarrow\xi=\zeta,\eta=\delta} = \underbrace{\underbrace{\underbrace{\overbrace{(\mathbf{A} \odot \mathbf{B})}^{\mu\nu\times\xi}}_{\Rightarrow\xi=\zeta}{}^\top}_{\xi\times\mu\nu} \underbrace{\underbrace{(\mathbf{C} \odot \mathbf{D})}_{\Rightarrow\eta=\delta}}_{\alpha\gamma\times\eta}}_{\Rightarrow\mu\nu=\alpha\gamma} \tag{37}$$

where "$\odot$" denotes the "columnwise Khatri-Rao product" and "$\circledast$" the scalar Hadamard product, see Table 2. We see that the constraints $\xi = \zeta$ and $\eta = \delta$ occur on both sides of (37). The constraint on the right-hand side (RHS) "$\mu\nu = \alpha\gamma$" is fulfilled with the left-hand side (LHS) constraints $\mu = \alpha$ and $\nu = \gamma$. We thus remain with

$$\mathbf{A}^\top\mathbf{C} \circledast \mathbf{B}^\top\mathbf{D} = (\mathbf{A}_{\mu\times\xi} \odot \mathbf{B}_{\nu\times\xi})^\top(\mathbf{C}_{\mu\times\eta} \odot \mathbf{D}_{\nu\times\eta}). \tag{38}$$

The constraints on the dimensions of the matrices imply that they may be regarded as sub-blocks of the same matrix,

$$\begin{bmatrix} \mathbf{A}_{\mu\times\xi} & \mathbf{C}_{\mu\times\eta} \\ \mathbf{B}_{\nu\times\xi} & \mathbf{D}_{\nu\times\eta} \end{bmatrix} \tag{39}$$

Proof of Identity A.4 and (37). On the LHS of (37), the (i, j)th scalar entry of $\mathbf{A}^\top\mathbf{C}$ is

$$[\mathbf{A}^\top\mathbf{C}]_{ij} = \sum_{k=1}^{\mu} a_{ki}c_{kj} = (\mathbf{a}_i)^\top\mathbf{c}_j \tag{40}$$

where a_{ki} is the (k, i)th scalar entry of $\mathbf{A}$ and $\mathbf{a}_i$ is the ith column vector of $\mathbf{A}$. Hence,

$$[\mathbf{A}^\top\mathbf{C} \circledast \mathbf{B}^\top\mathbf{D}]_{ij} = (\mathbf{a}_i)^\top\mathbf{c}_j \cdot (\mathbf{b}_i)^\top\mathbf{d}_j \tag{41}$$

On the RHS of (37),

$$(\mathbf{A}\odot\mathbf{B})^\top(\mathbf{C}\odot\mathbf{D}) = \begin{bmatrix} (\mathbf{a}_1\otimes\mathbf{b}_1)^\top \\ \vdots \\ (\mathbf{a}_K\otimes\mathbf{b}_K)^\top \end{bmatrix} \begin{bmatrix} \mathbf{c}_1\otimes\mathbf{d}_1 & \cdots & \mathbf{c}_K\otimes\mathbf{d}_K \end{bmatrix} \tag{42}$$

The (i, j)th block of (42) is

$$(\mathbf{a}_i\otimes\mathbf{b}_i)^\top(\mathbf{c}_j\otimes\mathbf{d}_j) = (\mathbf{a}_i^\top\mathbf{c}_j)\otimes(\mathbf{b}_i^\top\mathbf{d}_j) = (\mathbf{a}_i^\top\mathbf{c}_j)(\mathbf{b}_i^\top\mathbf{d}_j) = (41) \tag{43}$$

$\square$

Remark 4. Identity A.4 is a special case of Identity A.5.

Remark 5. Equation (37) is our extension to

$$\mathbf{A}^\top\mathbf{A} \circledast \mathbf{B}^\top\mathbf{B} = (\mathbf{A}\odot\mathbf{B})^\top(\mathbf{A}\odot\mathbf{B}), \tag{44}$$

see, e.g., [35, 36, 37, 38].

In order to solve the multidimensional case, we need to extend Identity A.4 (Equation (37)) and Identity A.2 (Equation (32)) to any block-partitions.

Identity A.5. Let $\mathbf{A}\in\mathbb{R}^{\mu\times a}$, $\mathbf{B}\in\mathbb{R}^{\phi\times b}$, $\mathbf{C}\in\mathbb{R}^{\mu\times c}$, $\mathbf{D}\in\mathbb{R}^{\nu\times d}$ be four matrices partitioned to column blocks as $\mathbf{A}_k\in\mathbb{R}^{\mu\times\alpha_k}$, $\mathbf{B}_k\in\mathbb{R}^{\mu\times\beta_k}$, $\mathbf{C}\in\mathbb{R}^{\nu\times\gamma_k}$, $\mathbf{D}\in\mathbb{R}^{\nu\times\delta_k}$, $\sum_{k=1}^K\alpha_k = a$, $\sum_{k=1}^K\beta_k = b$, $\sum_{k=1}^K\gamma_k = c$, $\sum_{k=1}^K\delta_k = d$. Then,

$$(\mathbf{A}\boxbar\mathbf{B})^\top(\mathbf{C}\boxbar\mathbf{D}) = \mathbf{A}^\top\mathbf{C}\boxplus\mathbf{B}^\top\mathbf{D} \tag{45}$$

Proof of Identity A.5 and (45).

$$\begin{aligned} \mathbf{A}_{\mu\times a}\boxbar\mathbf{B}_{\nu\times b} &= \left[\ \mathbf{A}_1 \mid \cdots \mid \mathbf{A}_K\ \right]\boxbar\left[\ \mathbf{B}_1 \mid \cdots \mid \mathbf{B}_K\ \right] \\ &= \left[\ \mathbf{A}_1\otimes\mathbf{B}_1 \mid \cdots \mid \underbrace{\mu\nu\times\alpha_k\beta_k}_{\mathbf{A}_k\otimes\mathbf{B}_k} \mid \cdots \mid \mathbf{A}_K\otimes\mathbf{B}_K\ \right] \in \mathbb{R}^{\mu\nu\times\sum_{k=1}^K\alpha_k\beta_k} \end{aligned} \tag{46}$$

Therefore,

$$\begin{aligned} (\mathbf{A}\boxbar\mathbf{B})^\top(\mathbf{C}\boxbar\mathbf{D}) &= \overbrace{\left[\ \mathbf{A}_1\otimes\mathbf{B}_1 \mid \cdots \mid \mathbf{A}_K\otimes\mathbf{B}_K\ \right]}^{\sum_{k=1}^K\alpha_k\beta_k\times\mu\nu}{}^\top \overbrace{\left[\ \mathbf{C}_1\otimes\mathbf{D}_1 \mid \cdots \mid \mathbf{C}_K\otimes\mathbf{D}_K\ \right]}^{\mu\nu\times\sum_{k=1}^K\gamma_k\delta_k} \\ &= \begin{bmatrix} (\mathbf{A}_1\otimes\mathbf{B}_1)^\top \\ \hline \vdots \\ \hline (\mathbf{A}_K\otimes\mathbf{B}_K)^\top \end{bmatrix} \left[\ \mathbf{C}_1\otimes\mathbf{D}_1 \mid \cdots \mid \mathbf{C}_K\otimes\mathbf{D}_K\ \right] \\ &= \begin{bmatrix} \mathbf{A}_1^\top\otimes\mathbf{B}_1^\top \\ \hline \vdots \\ \hline \mathbf{A}_K^\top\otimes\mathbf{B}_K^\top \end{bmatrix} \left[\ \mathbf{C}_1\otimes\mathbf{D}_1 \mid \cdots \mid \mathbf{C}_K\otimes\mathbf{D}_K\ \right] \\ &= \begin{bmatrix} \mathbf{A}_1^\top\mathbf{C}_1\otimes\mathbf{B}_1^\top\mathbf{D}_1 & \cdots & \mathbf{A}_1^\top\mathbf{C}_l\otimes\mathbf{B}_1^\top\mathbf{D}_l & \cdots & \mathbf{A}_1^\top\mathbf{C}_K\otimes\mathbf{B}_1^\top\mathbf{D}_K \\ \vdots & & & & \vdots \\ \mathbf{A}_k^\top\mathbf{C}_1\otimes\mathbf{B}_k^\top\mathbf{D}_1 & \cdots & \underbrace{\underbrace{\mathbf{A}_k^\top\mathbf{C}_l}_{[\mathbf{A}^\top\mathbf{C}]_{kl}\in\mathbb{R}^{\alpha_k\times\gamma_l}} \otimes \underbrace{\mathbf{B}_k^\top\mathbf{D}_l}_{[\mathbf{B}^\top\mathbf{D}]_{kl}\in\mathbb{R}^{\beta_k\times\delta_l}}}_{\alpha_k\beta_k\times\gamma_l\delta_l} & \cdots & \mathbf{A}_k^\top\mathbf{C}_K\otimes\mathbf{B}_k^\top\mathbf{D}_K \\ \vdots & & & & \vdots \\ \mathbf{A}_K^\top\mathbf{C}_1\otimes\mathbf{B}_K^\top\mathbf{D}_1 & \cdots & \mathbf{A}_K^\top\mathbf{C}_l\otimes\mathbf{B}_K^\top\mathbf{D}_l & \cdots & \mathbf{A}_K^\top\mathbf{C}_K\otimes\mathbf{B}_K^\top\mathbf{D}_K \end{bmatrix} \end{aligned} \tag{47}$$

In the above, we only need that $\mathbf{A}$ and $\mathbf{C}$ and $\mathbf{B}$ and $\mathbf{D}$ have the same number of rows (this can be formulated as $\begin{bmatrix}\mathbf{A} & \mathbf{C}\end{bmatrix}$ and $\begin{bmatrix}\mathbf{B} & \mathbf{D}\end{bmatrix}$, less constrained than (39)) and that all matrices are partitioned into the same number K of block columns. We do not need that $\alpha_k, \beta_k, \gamma_k, \delta_k$ be equal. The (k, l)th block of (47) is exactly the (k, l)th block of $\mathbf{A}^\top\mathbf{C}\boxplus\mathbf{B}^\top\mathbf{D}$. $\square$

B Proof of Lemma 1 and Lemma 2

Proof of Lemma 1 and Lemma 2 (Multiset analogue to Schur's first lemma). Lemma 1 states that K matrices $\mathbf{M}^{[k]}$ that commute with a set of strictly admissible (Definition 2.3) matrices $\{\mathbf{P}^{[k,l]}\}$, $k,l = 1,\ldots,K$,

$$\mathbf{M}^{[k]}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{M}^{[l]} \quad \forall k,l \tag{48}$$

are all equal to a constant multiple of the unit matrix, with the same constant value. Lemma 2 deals with cases in which some of the matrices in the set are zero. The proof that we propose here follows the same lines of the proof to Schur's original lemma in [4, Chapter 4], the difference arising from the fact that the original requirement from the commuting set to form a representation is relaxed, and of course the multiset scenario.

The first step in the proof is to show that each matrix $\mathbf{M}^{[k]}$ can be written as a sum of two Hermitian matrices. Let us write the conjugate transpose of (48) as

$$(\mathbf{P}^{[k,l]})^\dagger(\mathbf{M}^{[k]})^\dagger = (\mathbf{M}^{[l]})^\dagger(\mathbf{P}^{[k,l]})^\dagger \quad \forall k,l \tag{49}$$

Then, using either

$(\mathbf{P}^{[k,l]})^\dagger = (\mathbf{P}^{[k,l]})^{-1}$ if unitary,

$$(49) \quad \rightarrow \quad (\mathbf{P}^{[k,l]})^{-1}(\mathbf{M}^{[k]})^\dagger = (\mathbf{M}^{[l]})^\dagger(\mathbf{P}^{[k,l]})^{-1} \xrightarrow{\mathbf{P}^{[k,l]}\cdot\searrow\ \nearrow\cdot\mathbf{P}^{[k,l]}}$$

or

$(\mathbf{P}^{[k,l]})^\dagger = \mathbf{P}^{[l,k]}$ if symmetric ($\mathbf{P} = \mathbf{P}^\dagger$),

$$(49) \quad \rightarrow \quad \mathbf{P}^{[l,k]}(\mathbf{M}^{[k]})^\dagger = (\mathbf{M}^{[l]})^\dagger\mathbf{P}^{[l,k]} \xrightarrow{k\leftrightarrow l}$$

(49) rewrites as

$$(\mathbf{M}^{[k]})^\dagger\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}(\mathbf{M}^{[l]})^\dagger \quad \forall k,l \tag{50}$$

Hence, if K matrices $\mathbf{M}^{[k]}$ commute with $\mathbf{P}^{[k,l]}$, so do their conjugate transpose $(\mathbf{M}^{[k]})^\dagger$. Therefore, any linear combination of $\mathbf{M}^{[k]}$ and $(\mathbf{M}^{[k]})^\dagger$ commutes as well:

$$(a\mathbf{M}^{[k]} + b(\mathbf{M}^{[k]})^\dagger)\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}(a\mathbf{M}^{[l]} + b(\mathbf{M}^{[l]})^\dagger) \quad \forall a,b\in\mathbb{C}\ ,\ \forall k,l \tag{51}$$

In particular, $a = 1$, $b = 1$,

$$\underbrace{(\mathbf{M}^{[k]} + (\mathbf{M}^{[k]})^\dagger)}_{\mathbf{H}_1^{[k]}}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\underbrace{(\mathbf{M}^{[l]} + (\mathbf{M}^{[l]})^\dagger)}_{\mathbf{H}_1^{[l]}} \quad \forall k,l \tag{52}$$

and $a = i$, $b = -i$, where $i \triangleq \sqrt{-1}$,

$$\underbrace{i(\mathbf{M}^{[k]} - (\mathbf{M}^{[k]})^\dagger)}_{\mathbf{H}_2^{[k]}}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\ \underbrace{i(\mathbf{M}^{[l]} - (\mathbf{M}^{[l]})^\dagger)}_{\mathbf{H}_2^{[l]}} \quad \forall k,l \tag{53}$$

such that (easily verified)

$$\mathbf{M}^{[k]} = \frac{1}{2}(\mathbf{H}_1^{[k]} - i\mathbf{H}_2^{[k]})\ . \tag{54}$$

Note that $(\mathbf{H}_1^{[k]})^\dagger = \mathbf{H}_1^{[k]}$, $(\mathbf{H}_2^{[k]})^\dagger = \mathbf{H}_2^{[k]}$, i.e., Hermitian. The rest of the proof will show that, in the strict case, both $\mathbf{H}_1^{[k]}$ and $\mathbf{H}_2^{[k]}$ are proportional to the unit matrix up to two scalars $a,b\in\mathbb{R}$ such that $\mathbf{M}^{[k]} = \frac{1}{2}(a-ib)\mathbf{I}_{P^{[k]}}$. As we shall see, in the relaxed case, the situation is a bit more complicated. Hence, we shall now characterize the properties of the Hermitian matrices $\mathbf{H}^{[k]}$ that commute as

$$\mathbf{H}^{[k]}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{H}^{[l]} \quad \forall k,l\ . \tag{55}$$

The EVD of $\mathbf{H}^{[k]}$ is

$$\mathbf{H}^{[k]} = \mathbf{U}^{[k]}\mathbf{\Lambda}^{[k]}(\mathbf{U}^{[k]})^\dagger \tag{56}$$

where

$$\mathbf{\Lambda}^{[k]} \triangleq \begin{bmatrix} \lambda_1^{[k]} & & 0 \\ & \ddots & \\ 0 & & \lambda_{P^{[k]}}^{[k]} \end{bmatrix}, \quad \lambda_p^{[k]} \in \mathbb{R}$$

Substituting (56) in (55),

$$\mathbf{U}^{[k]}\mathbf{\Lambda}^{[k]}(\mathbf{U}^{[k]})^\dagger \mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{U}^{[l]}\mathbf{\Lambda}^{[l]}(\mathbf{U}^{[l]})^\dagger \quad \forall k,l$$

and isolating the diagonal terms,

$$\mathbf{\Lambda}^{[k]}\underbrace{\mathbf{U}^{[k]\dagger}\mathbf{P}^{[k,l]}\mathbf{U}^{[l]}}_{\mathbf{C}^{[k,l]}} = \mathbf{U}^{\dagger[k]}\mathbf{P}^{[k,l]}\mathbf{U}^{[l]}\mathbf{\Lambda}^{[l]} \quad \text{where} \quad \mathbf{C}^{[k,l]} \triangleq \mathbf{U}^{[k]\dagger}\mathbf{P}^{[k,l]}\mathbf{U}^{[l]} \quad \forall k,l \tag{57}$$

we obtain

$$\mathbf{\Lambda}^{[k]}\mathbf{C}^{[k,l]} = \mathbf{C}^{[k,l]}\mathbf{\Lambda}^{[l]} \quad \forall k,l \tag{58}$$

Equation (58) looks similar to (48); however, due to the diagonal structure of $\mathbf{\Lambda}^{[k]}$, (58) is a simpler problem. The (α,β)th scalar element of (58) is

$$[\mathbf{\Lambda}^{[k]}\mathbf{C}^{[k,l]}]_{\alpha\beta} = \sum_{p=1}^{P}[\mathbf{\Lambda}^{[k]}]_{\alpha p}c_{p\beta}^{[k,l]} = [\mathbf{\Lambda}^{[k]}]_{\alpha\alpha}c_{\alpha\beta}^{[k,l]}$$
$$[\mathbf{C}^{[k,l]}\mathbf{\Lambda}^{[l]}]_{\alpha\beta} = \sum_{p=1}^{P}c_{\alpha p}^{[k,l]}[\mathbf{\Lambda}^{[l]}]_{p\beta} = c_{\alpha\beta}^{[k,l]}[\mathbf{\Lambda}^{[l]}]_{\beta\beta}$$

Therefore, (58) rewrites as

$$\lambda_\alpha^{[k]}c_{\alpha\beta}^{[k,l]} = c_{\alpha\beta}^{[k,l]}\lambda_\beta^{[l]} \quad \forall k,l$$

where $\alpha = 1,\ldots,P^{[k]}$, $\beta = 1,\ldots,P^{[l]}$. Changing sides,

$$(\lambda_\alpha^{[k]} - \lambda_\beta^{[l]})c_{\alpha\beta}^{[k,l]} = 0 \quad \forall k,l \tag{59}$$

We now set out to characterize the non-trivial solutions to (59).

Case B.1: All diagonal elements of $\mathbf{\Lambda}^{[k]}$ equal (and non-zero).

$$\mathbf{\Lambda}^{[k]} = \lambda\mathbf{I}_{P^{[k]}} \quad \forall k \quad \lambda \neq 0$$

Looking at (59), these values do not impose any constraints on $\mathbf{C}$.

► Going back to the EVD (56), $\mathbf{\Lambda}^{[k]} = \lambda\mathbf{I}_{P^{[k]}}$ implies

$$\mathbf{H}^{[k]} = \mathbf{U}^{[k]}\lambda\mathbf{I}_{P^{[k]}}\mathbf{U}^{[k]\dagger} = \lambda\mathbf{I}_{P^{[k]}} \quad \forall k \tag{60}$$

In order to reconstruct $\mathbf{M}^{[k]}$, we associate two such values, $\lambda_1, \lambda_2 \in \mathbb{R}$, with $\mathbf{H}_1^{[k]}, \mathbf{H}_2^{[k]}$, $\forall k$, respectively. Substituting this result in (54) implies that $\mathbf{M}^{[k]} = \mu\mathbf{I}_{P^{[k]}}$ where $\mu = \frac{1}{2}(\lambda_1 - i\lambda_2) \in \mathbb{C}$.

Case B.2: All diagonal elements of $\mathbf{\Lambda}^{[k]}$ distinct.

$$\underbrace{\lambda_\alpha^{[k]} \neq \lambda_\beta^{[l]}}_{\text{one may be zero}} \quad \text{if} \quad (\alpha,k) \neq (\beta,l) \tag{61}$$

Equation (59) implies that

$$(\lambda_\alpha^{[k]} - \lambda_\beta^{[l]}) \neq 0 \quad \Rightarrow \quad c_{\alpha\beta}^{[k,l]} = 0 \quad \forall(\alpha,k) \neq (\beta,l)$$

Hence, non-zero $c_{\alpha\beta}^{[k,l]}$ remain only at $c_{\alpha\alpha}^{[k,k]}$, inducing a structure

$$\mathbf{C} = \text{diag}\{\mathbf{C}\} \quad \Leftrightarrow \quad \mathbf{C}^{[k,l]} = \begin{cases} \mathbf{0} & k \neq l \\ \text{diag}\{\mathbf{C}^{[k,k]}\} & k = l \end{cases} \tag{62}$$

which is not strictly admissible yet admissible in the relaxed sense.

▶ Going back to the EVD (56), distinct eigenvalues imply that $\mathbf{H}^{[k]} = \mathbf{U}^{[k]}\mathbf{\Lambda}^{[k]}(\mathbf{U}^{[k]})^\dagger$ is an arbitrary Hermitian matrix. For each $\mathbf{H}_1^{[k]}$ and $\mathbf{H}_2^{[k]}$ we can associate different arbitrary $\mathbf{\Lambda}_1^{[k]}, \mathbf{\Lambda}_2^{[k]}$ with real values, according to (61). Substituting in (54) yields

$$\mathbf{M}^{[k]} = \frac{1}{2}(\mathbf{U}^{[k]}\mathbf{\Lambda}_1^{[k]}(\mathbf{U}^{[k]})^\dagger - i\mathbf{U}^{[k]}\mathbf{\Lambda}_2^{[k]}(\mathbf{U}^{[k]})^\dagger) = \frac{1}{2}(\mathbf{U}^{[k]}(\mathbf{\Lambda}_1^{[k]} - i\mathbf{\Lambda}_2^{[k]})(\mathbf{U}^{[k]})^\dagger) = \frac{1}{2}(\mathbf{U}^{[k]}\mathbf{\Lambda}_3^{[k]}(\mathbf{U}^{[k]})^\dagger) \tag{63}$$

which is a Hermitian matrix. We now explain how to find $\mathbf{U}^{[k]}$. It follows from (57) and (62) that

$$\mathbf{P}^{[k,l]} = \begin{cases} \mathbf{U}^{[k]} \operatorname{diag}\{\mathbf{C}^{[k,k]}\}(\mathbf{U}^{[k]})^\dagger & k = l \\ \mathbf{0} & k \neq l \end{cases} \tag{64}$$

Hence, $\operatorname{diag}\{\mathbf{C}^{[k,k]}\}$ can be interpreted as the matrix whose diagonal consists of the eigenvalues of $\mathbf{P}^{[k,k]}$. Therefore, $\mathbf{U}^{[k]}$ can be obtained from the EVD of $\mathbf{P}^{[k,k]}$. We conclude that for data with structure (64), one can always find an arbitrary set of Hermitian matrices, not all zero, such that (48) holds.

Case B.3: In $D < K$ mixtures, the diagonal elements of $\mathbf{\Lambda}^{[k]}$ are equal, and different from the rest. W.l.o.g., we assume that those D mixtures are in indices $\mathcal{D} = \{1, \ldots, D\}$. Then,

$$D < K: \quad \underbrace{\overbrace{\lambda_\alpha^{[1]} = \ldots = \lambda_\beta^{[D]} = \lambda}^{\mathbf{\Lambda}^{[k\in\mathcal{D}]} = \lambda\mathbf{I}_{P^{[k]}}}}_{\neq 0} \neq \lambda_\gamma^{[k\notin\mathcal{D}]}$$

Similarly to the analysis in Case B.1, (59) does not impose any constraints on $\mathbf{C}^{[k\in\mathcal{D},l\in\mathcal{D}]}$ nor $\mathbf{C}^{[k\notin\mathcal{D},l\notin\mathcal{D}]}$. However, it does impose $\mathbf{C}^{[k\in\mathcal{D},l\notin\mathcal{D}]} = \mathbf{0} = \mathbf{C}^{[k\notin\mathcal{D},l\in\mathcal{D}]}$. These constraints can be reformulated as

$$\mathbf{C} = \left[\begin{array}{c|c} \mathbf{C}^{[1:D,1:D]} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array}\right] \tag{65}$$

where the top right block of zeros has size $\left(\sum_{k=1}^{D} P^{[k]}\right) \times \left(\sum_{k=D+1}^{K} P^{[k]}\right)$ and the bottom left is its transpose. This case is not strictly admissible yet admissible in the relaxed sense.

▶ The analysis follows the same lines as Case B.1 ($D = K$), where now we consider only the upper $D \times D$ block of $\mathbf{C}$. The EVD (56) implies that

$$\mathbf{H}^{[k]} = \mathbf{U}^{[k]}\lambda\mathbf{I}_{P^{[k]}}(\mathbf{U}^{[k]})^\dagger = \lambda\mathbf{I}_{P^{[k]}} \quad , \ k \in \mathcal{D}$$

We can write (54) as $\mathbf{M}^{[k\in\mathcal{D}]} = \mu\mathbf{I}_{P^{[k]}}$, where $\mu = \frac{1}{2}(\lambda_1 - i\lambda_2)$. Without further knowledge of $\mathbf{C}^{[D+1:K,D+1:K]}$, we can take $\mathbf{M}^{[k\notin\mathcal{D}]} = \mathbf{0}$.

Case B.4: In $D < K$ mixtures, the diagonal elements of $\mathbf{\Lambda}^{[k]}$ are all distinct, and different from the rest. W.l.o.g., we assume that those D mixtures are in indices $k \in \mathcal{D} = \{1, \ldots, D\}$. Then,

$$D < K: \quad \underbrace{\lambda_\alpha^{[k\in\mathcal{D}]} \neq \lambda_\beta^{[l\in\mathcal{D}]}}_{\neq 0} \neq \text{ all others} \quad \forall(\alpha, k) \neq (\beta, l) \tag{66}$$

Equation (59) implies that for any $k, l \in \mathcal{D}$,

$$(\lambda_\alpha^{[k]} - \lambda_\beta^{[l]}) \neq 0 \quad \Rightarrow \quad c_{\alpha\beta}^{[k,l]} = 0 \quad \forall(\alpha, k) \neq (\beta, l)$$

Hence, non-zero $c_{\alpha\beta}^{[k,l]}$ remain only at $c_{\alpha\alpha}^{[k,k]}$, $k \in \mathcal{D}$. These constraints can be reformulated as

$$\mathbf{C} = \left[\begin{array}{c|c} \operatorname{diag}\{\mathbf{C}^{[1:D,1:D]}\} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array}\right] \quad \Leftrightarrow \quad \mathbf{C}^{[k,l]} = \begin{cases} \mathbf{0} & k, l \in \mathcal{D}\ ,\ k \neq l \\ \operatorname{diag}\{\mathbf{C}^{[k,k]}\} & k \in \mathcal{D} \\ \mathbf{C}^{[k,l]} \text{ (unconstrained)} & k, l \notin \mathcal{D} \end{cases} \tag{67}$$

This structure is not strictly admissible yet admissible in the relaxed sense, as can be verified from (57).

▶ The analysis follows the same lines as Case B.2 ($D = K$), where now we consider only the upper $D \times D$ block of $\mathbf{C}$. It follows that for this structure, one can always use (63) to construct D Hermitian $\mathbf{M}^{[k]} \neq \mathbf{0}$, $k \in \mathcal{D}$, whose unitary matrices are those of the EVD of $\mathbf{P}^{[k,k]}$ and the diagonal values arbitrary $\in \mathbb{C}$. Without further knowledge of $\mathbf{C}^{[D+1:K,D+1:K]}$, we can take $\mathbf{M}^{[k\notin\mathcal{D}]} = \mathbf{0}$.

Case B.5: In all mixtures, $G^{[k]} < P^{[k]}$ diagonal elements of $\mathbf{\Lambda}^{[k]}$ are identical, and different from the rest. W.l.o.g., $\mathcal{G}^{[k]} = \{1, \ldots, G^{[k]}\}$ $\forall k$ and $|\mathcal{G}^{[k]}| < P^{[k]}$. Then,

$$\underbrace{\lambda^{[k]}_{\alpha \in \mathcal{G}^{[k]}}}_{\neq 0} = \lambda \neq \text{ all others}$$

This scenario can be written as

$$\mathbf{\Lambda}^{[k]} = \left[\begin{array}{c|c} \lambda \mathbf{I}_{G^{[k]}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Lambda}'^{[k]} \end{array}\right] \quad \forall k$$

where in $\mathbf{\Lambda}'^{[k]} \triangleq \text{diag}\{\lambda^{[k]}_{G^{[k]}+1}, \ldots, \lambda^{[k]}_{P^{[k]}}\}$, no element is equal to λ. In order to simplify notations, in what follows, we omit the superscript of $\mathcal{G}$ whenever it is implicit, e.g., $\lambda^{[k]}_{\alpha \in \mathcal{G}} \triangleq \lambda^{[k]}_{\alpha \in \mathcal{G}^{[k]}}$. For any pair (k, l), (59) implies that

$$\begin{aligned}
\underbrace{(\lambda^{[k]}_{\alpha \in \mathcal{G}} - \lambda^{[l]}_{\beta \in \mathcal{G}})}_{0} c^{[k,l]}_{\alpha \in \mathcal{G}, \beta \in \mathcal{G}} = 0 \quad &\Rightarrow \quad \text{no constraints on } c^{[k,l]}_{\alpha \in \mathcal{G}, \beta \in \mathcal{G}} \\
\underbrace{(\lambda^{[k]}_{\alpha \in \mathcal{G}} - \lambda^{[l]}_{\beta \notin \mathcal{G}})}_{\neq 0} c^{[k,l]}_{\alpha \in \mathcal{G}, \beta \notin \mathcal{G}} = 0 \quad &\Rightarrow \quad c^{[k,l]}_{\alpha \in \mathcal{G}, \beta \notin \mathcal{G}} = 0 \\
\underbrace{(\lambda^{[k]}_{\alpha \notin \mathcal{G}} - \lambda^{[l]}_{\beta \in \mathcal{G}})}_{\neq 0} c^{[k,l]}_{\alpha \notin \mathcal{G}, \beta \in \mathcal{G}} = 0 \quad &\Rightarrow \quad c^{[k,l]}_{\alpha \notin \mathcal{G}, \beta \in \mathcal{G}} = 0 \\
\underbrace{(\lambda^{[k]}_{\alpha \notin \mathcal{G}} - \lambda^{[l]}_{\beta \notin \mathcal{G}})}_{\neq 0} c^{[k,l]}_{\alpha \notin \mathcal{G}, \beta \notin \mathcal{G}} = 0 \quad &\Rightarrow \quad \text{no information about } c^{[k,l]}_{\alpha \notin \mathcal{G}, \beta \notin \mathcal{G}}
\end{aligned}$$

These constraints correspond to $\mathbf{C}^{[k,l]}$ that have the general form

$$\mathbf{C}^{[k,l]} = \left[\begin{array}{c|c} \square_{G^{[k]} \times G^{[l]}} & \mathbf{0}_{G^{[k]} \times (P^{[l]} - G^{[l]})} \\ \hline \mathbf{0}_{(P^{[k]} - G^{[k]}) \times G^{[l]}} & \square_{(P^{[k]} - G^{[k]}) \times (P^{[l]} - G^{[l]})} \end{array}\right]$$

where $\square$ denotes a matrix with arbitrary values. In this case, the set is reducible in the generalized sense (Definition 2.1) and thus the solution is not admissible.

Case B.6: In $1 \leq D < K$ mixtures, $G^{[k]} < P^{[k]}$ diagonal elements of $\mathbf{\Lambda}^{[k]}$ are identical, and different from the rest, excluding[1] "$D = 1$ and $G^{[1]} = 1$". W.l.o.g., $\mathcal{G}^{[k]} = \{1, \ldots, G^{[k]}\}$ and $|\mathcal{G}^{[k]}| < P^{[k]}$ $\forall k$. Then,

$$D < K: \quad \underbrace{\lambda^{[k \in \mathcal{D}]}_{\alpha \in \mathcal{G}^{[k]}}}_{\neq 0} = \lambda \neq \text{ all others}$$

This scenario can be written as

$$\mathbf{\Lambda}^{[k \in \mathcal{D}]} = \left[\begin{array}{c|c} \lambda \mathbf{I}_{G^{[k]}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Lambda}'^{[k]} \end{array}\right]$$

where in $\mathbf{\Lambda}'^{[k \in \mathcal{D}]} \triangleq \text{diag}\{\lambda^{[k]}_{G^{[k]}+1}, \ldots, \lambda^{[k]}_{P^{[k]}}\}$, as well as in all $\mathbf{\Lambda}^{[k \notin \mathcal{D}]}$, no element is equal to λ. As in Case B.5 ($D = K$), we omit the superscript of $\mathcal{G}^{[k]}$ whenever it is implicit. Within the first D mixtures, i.e., $k, l \in \mathcal{D}$, (59) implies

$$\begin{aligned}
\underbrace{(\lambda^{[k \in \mathcal{D}]}_{\alpha \in \mathcal{G}} - \lambda^{[l \in \mathcal{D}]}_{\beta \in \mathcal{G}})}_{0} c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \in \mathcal{G}, \beta \in \mathcal{G}} = 0 \quad &\Rightarrow \quad \text{no constraints on } c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \in \mathcal{G}, \beta \in \mathcal{G}} \\
\underbrace{(\lambda^{[k \in \mathcal{D}]}_{\alpha \in \mathcal{G}} - \lambda^{[l \in \mathcal{D}]}_{\beta \notin \mathcal{G}})}_{\neq 0} c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \in \mathcal{G}, \beta \notin \mathcal{G}} = 0 \quad &\Rightarrow \quad c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \in \mathcal{G}, \beta \notin \mathcal{G}} = 0 \\
\underbrace{(\lambda^{[k \in \mathcal{D}]}_{\alpha \notin \mathcal{G}} - \lambda^{[l \in \mathcal{D}]}_{\beta \in \mathcal{G}})}_{\neq 0} c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \notin \mathcal{G}, \beta \in \mathcal{G}} = 0 \quad &\Rightarrow \quad c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \notin \mathcal{G}, \beta \in \mathcal{G}} = 0 \\
\underbrace{(\lambda^{[k \in \mathcal{D}]}_{\alpha \notin \mathcal{G}} - \lambda^{[l \in \mathcal{D}]}_{\beta \notin \mathcal{G}})}_{\neq 0} c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \notin \mathcal{G}, \beta \notin \mathcal{G}} = 0 \quad &\Rightarrow \quad \text{no information about } c^{[k \in \mathcal{D}, l \in \mathcal{D}]}_{\alpha \notin \mathcal{G}, \beta \notin \mathcal{G}}
\end{aligned}$$

[1]The case "$D = 1$ and $G^{[1]} = 1$", that is, "in one dataset, one value is different from the rest", is not well-defined: w.l.o.g., if $\lambda^{[1]}_1$ is different from the rest, it is possible that all $\lambda^{[k]}_\alpha$ are distinct, hence, Case B.2. However, there may also exist at least two other eigenvalues that are identical, i.e., Case B.3 or B.6.

For $k \in \mathcal{D}$ and $l \notin \mathcal{D}$, and similarly for $k \notin \mathcal{D}$ and $l \in \mathcal{D}$

$$\begin{aligned}
&\underbrace{(\lambda_{\alpha\in\mathcal{G}}^{[k\in\mathcal{D}]} - \lambda_{\forall\beta}^{[l\notin\mathcal{D}]})}_{\neq 0} c_{\alpha\in\mathcal{G},\forall\beta}^{[k\in\mathcal{D},l\notin\mathcal{D}]} = 0 \quad \Rightarrow \quad c_{\alpha\in\mathcal{G},\forall\beta}^{[k\in\mathcal{D},l\notin\mathcal{D}]} = 0 \\
&\underbrace{(\lambda_{\alpha\notin\mathcal{G}}^{[k\in\mathcal{D}]} - \lambda_{\forall\beta}^{[l\notin\mathcal{D}]})}_{?} c_{\alpha\notin\mathcal{G},\forall\beta}^{[k\in\mathcal{D},l\notin\mathcal{D}]} = 0 \quad \Rightarrow \quad \text{no information about } c_{\alpha\notin\mathcal{G},\forall\beta}^{[k\in\mathcal{D},l\notin\mathcal{D}]}
\end{aligned}$$

For $k, l \notin \mathcal{D}$,

$$(\underbrace{\lambda_{\alpha}^{[k\notin\mathcal{D}]}}_{?} - \underbrace{\lambda_{\beta}^{[l\notin\mathcal{D}]}}_{?}) c_{\alpha,\beta}^{[k\notin\mathcal{D},l\notin\mathcal{D}]} = 0 \quad \Rightarrow \quad \text{no information about } c_{\alpha,\beta}^{[k\notin\mathcal{D},l\notin\mathcal{D}]}$$

If the values of λ outside $\mathcal{G}$ are zero, that is, $\lambda_{G^{[k]}<\alpha\leq P^{[k]}}^{[k]} = 0 \; \forall k$, then "no information" becomes "unconstrained". These constraints impose the following structure on $\mathbf{C}$:

$$\begin{aligned}
\mathbf{C}^{[k\in\mathcal{D},l\in\mathcal{D}]} &= \left[\begin{array}{c|c} \square_{G^{[k]}\times G^{[l]}} & \mathbf{0}_{G^{[k]}\times(P^{[l]}-G^{[l]})} \\ \hline \mathbf{0}_{(P^{[k]}-G^{[k]})\times G^{[l]}} & \square_{(P^{[k]}-G^{[k]})\times(P^{[l]}-G^{[l]})} \end{array}\right] \\
\mathbf{C}^{[k\in\mathcal{D},l\notin\mathcal{D}]} &= \left[\begin{array}{c} \mathbf{0}_{G^{[k]}\times P^{[l]}} \\ \hline \square_{(P^{[k]}-G^{[k]})\times P^{[l]}} \end{array}\right] = (\mathbf{C}^{[k\notin\mathcal{D},l\in\mathcal{D}]})^{\dagger} \\
\mathbf{C}^{[k\notin\mathcal{D},l\notin\mathcal{D}]} &= \square_{P^{[k]}\times P^{[l]}}
\end{aligned}$$

If at least one value of $\square_{(P^{[k]}-G^{[k]})\times P^{[l]}}$ in at least one $\mathbf{C}^{[k\in\mathcal{D},l\notin\mathcal{D}]}$ is not zero, then this structure is not admissible because there exists at least one $\mathbf{C}^{[k\in\mathcal{D},l\notin\mathcal{D}]}$ which is not full-rank yet no zero. If all $\square_{(P^{[k]}-G^{[k]})\times P^{[l]}}$ in all $\mathbf{C}^{[k\in\mathcal{D},l\notin\mathcal{D}]}$ are zero, then this structure is not admissible because it is reducible in the generalized sense. Either way, we conclude that this case is always not admissible.

The following roadmap ("in D mixtures, $G^{[k]}$ diagonal elements are identical, and different from the rest") summarizes our analysis.

	$G^{[k]} = P^{[k]}$	$G^{[k]} = 0$	$1 \leq G^{[k]} < P^{[k]}$ 1
$D = K$	Case B.1	Case B.2	Case B.5
$D < K$	Case B.3	Case B.4	Case B.6

Case B.1 is the only admissible solution in the strict scenario and thus proves Lemma 1. When the admissibility conditions are relaxed, Cases B.2, B.3 and B.4 provide admissible solutions. In order to obtain further insights, note that the link between $\mathbf{C}$ and $\mathbf{P}$, established by (57), implies

$$\mathbf{P}^{[k,l]} = \begin{cases} \mathbf{0} & \mathbf{C}^{[k,l]} = \mathbf{0} \\ \mathbf{U}^{[k]} \operatorname{diag}\{\mathbf{C}^{[k,k]}\}(\mathbf{U}^{[k]})^{\dagger} & \operatorname{diag}\{\mathbf{C}^{[k,k]}\} \end{cases}$$

Hence, in terms of $\mathbf{P}$, there is no difference between $\mathbf{C}^{[k,k]}$ and $\operatorname{diag}\{\mathbf{C}^{[k,k]}\}$: it is always a full matrix. Furthermore, in terms of structure of $\mathbf{P}$, there is no difference between (65) and (67) when $D = 1$. The above cases can now be summarized as follows:

Scenario B.1.

$$2 \leq D \leq K \;:\; \mathbf{\Lambda}^{[k\in\mathcal{D}]} = \lambda\mathbf{I}_{P^{[k]}} \,,\; \lambda \neq \lambda_{\gamma}^{[k\notin\mathcal{D}]} \quad , \quad \lambda \in \mathbb{R} \,,\; \lambda \neq 0$$

i.e., in the first D datasets, all eigenvalues identical and non-zero, and different from the rest. Such values are associated with admissible data of the form

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{P}^{[1:D,1:D]} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{P}^{[D+1:K,D+1:K]} \end{array}\right] \tag{68}$$

where the off-diagonal block of zeros on the top right has size $\left(\sum_{k=1}^{D} P^{[k]}\right) \times \left(\sum_{k=D+1}^{K} P^{[k]}\right)$, and its transpose on the bottom left. The generalized commutation relation (48) holds for any $\mathbf{M}^{[k\in\mathcal{D}]} = \mu\mathbf{I}_{P^{[k]}}$, where $\mu \in \mathbb{C}$, and $\mathbf{M}^{[k\notin\mathcal{D}]} = \mathbf{0}$.

Scenario B.2.

$$1 \le D \le K \ : \ \underbrace{\overbrace{\lambda_\alpha^{[1]} \neq \ldots \neq \lambda_\beta^{[D]}}^{\text{all eigenvalues distinct}}}_{\text{one may be zero}} \neq \lambda_\gamma^{[k>D]} \quad , \ \lambda_\alpha^{[k]} \in \mathbb{R}$$

i.e., in the first D datasets, all eigenvalues distinct, at most one zero, and different from the rest. Such values are associated with admissible data of the form

$$\mathbf{P} = \left[\begin{array}{ccc|c} \mathbf{P}^{[1,1]} & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \mathbf{P}^{[D,D]} & \\ \hline & \mathbf{0} & & \mathbf{P}^{[D+1:K,D+1:K]} \end{array}\right] \tag{69}$$

where the off-diagonal block of zeros on the top right has size $\left(\sum_{k=1}^{D} P^{[k]}\right) \times \left(\sum_{k=D+1}^{K} P^{[k]}\right)$, and its transpose on the bottom left. The generalized commutation relation (48) holds for any D Hermitian matrices $\mathbf{M}^{[k\in\mathcal{D}]} = \mathbf{U}^{[k]} \operatorname{diag}\{\mu_1^{[k]}, \ldots, \mu_{P^{[k]}}^{[k]}\}(\mathbf{U}^{[k]})^\dagger$, $\mu_p^{[k]} \in \mathbb{C}$ where $\mathbf{U}^{[k]}$ are the unitary matrices of the EVD of $\mathbf{P}^{[k,k]}$, and $\mathbf{M}^{[k\notin\mathcal{D}]} = \mathbf{0}$.

Whenever $D < K$, we did not specify what to do with $\mathbf{M}^{[k\notin\mathcal{D}]}$. Depending on the values at $k \notin \mathcal{D}$, and using the fact that the ordering of the datasets is arbitrary, the same considerations should be applied to datasets $k \notin \mathcal{D}$. It is thus possible that a solution with non-zero $\mathbf{M}^{[k\notin\mathcal{D}]}$ exists. This leads us to Lemma 2. □

C Proof of Lemma 3 and Lemma 4

Proof of Lemma 3 and Lemma 4. Our proof follows the same lines as the proof to Schur's second Lemma in [4, Chapter 4.2] and [39, Lemma A.4], for example, with the necessary adaptations for data that are not a representation, possibly with zero values, and of course the multiset setup. We have already used this approach for the identifiability analysis of the one-dimensional case [10].

In a first step, we multiply (1) on the left with $\mathbf{L}^{[k]\dagger}$, and, separately, on the right with $\mathbf{L}^{[l]\dagger}$:

$$\mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]}\mathbf{L}^{[l]} = \mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} \tag{70a}$$

$$\mathbf{L}^{[k]}\mathbf{R}^{[k,l]}\mathbf{L}^{[l]\dagger} = \mathbf{P}^{[k,l]}\mathbf{L}^{[l]}\mathbf{L}^{[l]\dagger} \tag{70b}$$

From here, there are two options.

Unitary: If all $\mathbf{P}^{[k,l]}$ unitary, taking the transpose of (1),

$$\mathbf{R}^{[k,l]\dagger}\mathbf{L}^{[k]\dagger} = \mathbf{L}^{[l]\dagger}\mathbf{P}^{[k,l]\dagger} \quad \xrightarrow{(\mathbf{P}^{[k,l]})^\dagger = (\mathbf{P}^{[k,l]})^{-1}} \quad (\mathbf{R}^{[k,l]})^{-1}\mathbf{L}^{[k]\dagger} = \mathbf{L}^{[l]\dagger}(\mathbf{P}^{[k,l]})^{-1}$$

Multiplying on the left with $\mathbf{R}^{[k,l]}$ and on the right with $\mathbf{P}^{[k,l]}$,

$$\mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]} = \mathbf{R}^{[k,l]}\mathbf{L}^{[l]\dagger}$$

Multiplying on the left with $\mathbf{L}^{[k]}$, and, separately, on the right with $\mathbf{L}^{[l]}$,

$$\mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]}\mathbf{L}^{[l]} = \mathbf{R}^{[k,l]}\mathbf{L}^{[l]\dagger}\mathbf{L}^{[l]} \tag{71a}$$

$$\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]} = \mathbf{L}^{[k]}\mathbf{R}^{[k,l]}\mathbf{L}^{[l]\dagger} \tag{71b}$$

Since the RHS of (71) is equal to the LHS of (70), we can write

$$\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{L}^{[l]}\mathbf{L}^{[l]\dagger} \tag{72a}$$

$$\mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{R}^{[k,l]}\mathbf{L}^{[l]\dagger}\mathbf{L}^{[l]} \tag{72b}$$

Symmetric: If $(\mathbf{P}^{[k,l]})^\dagger = (\mathbf{P}^{[l,k]})$, then the LHS of (70) is symmetric, and the same for its RHS. Hence, (72).

The importance of this step is that it decouples the original problem (1) into two simpler and equivalent problems (72), each with a single dataset. Thus, the set of $P^{[k]} \times P^{[k]}$ matrices $\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger}$ commutes with all the matrices of the irreducible set $\{\mathbf{P}^{[k,l]}\}$, and similarly for the set of $R^{[k]} \times R^{[k]}$ matrices $\mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]}$ and the irreducible set $\{\mathbf{R}^{[k,l]}\}$. ♠[2]

[2] Here is the bifurcation point between strict and relaxed

Proof of Lemma 3 Applying Lemma 1 to (72), the two sets of Gramian matrices $\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger}$ and $\mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]}$ must be a multiple of the unit matrix with a real constant factor:

$$\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger} = \mu_P \mathbf{I}_{P^{[k]}} \quad , \quad \mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]} = \mu_R \mathbf{I}_{R^{[k]}} \quad , \quad \mu_P, \mu_R \in \mathbb{R}_{\geq 0} \quad \forall k \tag{73}$$

Consider now the following cases.

Case C.1: $P^{[k]} = R^{[k]}$ $\forall k$**, strict.** If $\mu_P \neq 0$ (i.e., $\mu_P \in \mathbb{R}_{>0}$) and $P^{[k]} = R^{[k]}$, then (73) implies that $\mathbf{L}^{[k]}$ is invertible. This implies $\mu_P = \mu_R \triangleq \mu$ and

$$(\frac{1}{\sqrt{\mu}}\mathbf{L}^{[k]})^{-1} = (\frac{1}{\sqrt{\mu}}\mathbf{L}^{[k]})^{\dagger} \quad \forall k$$

The same would obviously result if we begin with $\mu_R \neq 0$. By definition, $\frac{1}{\sqrt{\mu}}\mathbf{L}^{[k]} \triangleq \mathbf{O}^{[k]}$ are unitary matrices. Substituting $\mathbf{L}^{[k]} = \sqrt{\mu}\mathbf{O}^{[k]}$ in (1),

$$\sqrt{\mu}\mathbf{O}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]}\sqrt{\mu}\mathbf{O}^{[l]} \quad \forall k, l \tag{74}$$

Eliminating the constant and changing sides,

$$\mathbf{P}^{[k,l]} = \mathbf{O}^{[k]}\mathbf{R}^{[k,l]}\mathbf{O}^{[l]\dagger} \quad \forall k, l \tag{75}$$

which is a generalized similarity transformation (an equivalence relation, Definition 2.2).

If $\mu_P = 0$, then $\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger} = \mathbf{0}_{P^{[k]}\times P^{[k]}}$. Its (i,j)th entry is $[\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger}]_{ij} = \sum_{p=1}^{P^{[k]}} L_{ip}L_{jp}^{*} = 0$. In particular, $[\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger}]_{ii} = \sum_{p=1}^{P^{[k]}} |L_{ip}|^2 = 0$ which implies $L_{ip} = 0$ $\forall i, p$, i.e., $\mathbf{L}^{[k]} = \mathbf{0}_{P^{[k]}\times P^{[k]}}$ $\forall k$.

Case C.2: $P^{[k]} \neq R^{[k]}$ **for at least one** k**, strict.** W.l.o.g., let $P^{[k]} > R^{[k]}$ for some k. Then, we can extend the corresponding $\mathbf{L}^{[k]}$ to a rectangular $P^{[k]} \times P^{[k]}$ by adding a $P^{[k]} \times (P^{[k]} - R^{[k]})$ matrix of zeros to its right, $\mathbf{N}^{[k]} \triangleq \left[\begin{array}{c|c} \mathbf{L}^{[k]} & \mathbf{0}_{P^{[k]}\times(P^{[k]}-R^{[k]})} \end{array}\right]$. Then,

$$\mathbf{N}^{[k]}\mathbf{N}^{[k]\dagger} = \left[\begin{array}{c|c} \mathbf{L}^{[k]} & \mathbf{0}_{P^{[k]}\times(P^{[k]}-R^{[k]})} \end{array}\right] \left[\begin{array}{c} \mathbf{L}^{[k]\dagger} \\ \hline \mathbf{0}_{(P^{[k]}-R^{[k]})\times P^{[k]}} \end{array}\right] = \mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger} = \mu_P \mathbf{I}_{P^{[k]}}$$

where the last equality is by (73). The determinant of $\mathbf{N}^{[k]}$ is zero, hence $\det(\mathbf{N}^{[k]}\mathbf{N}^{[k]\dagger}) = \det(\mathbf{N}^{[k]})\det(\mathbf{N}^{[k]\dagger}) = 0 = (\mu_P)^{P^{[k]}}$. Hence, $\mu_P = 0$, which implies $\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger} = \mathbf{0}_{P^{[k]}\times P^{[k]}}$. Proceeding as in Case C.1, we conclude that $\mathbf{L}^{[k]} = \mathbf{0}_{P^{[k]}\times P^{[k]}}$ $\forall k$.

The analysis of the strict case is summarized in the following roadmap:

	Strict
$P^{[k]} = R^{[k]}$ $\forall k$	Case C.1
$P^{[k]} \neq R^{[k]}$ for some k	Case C.2

Proof of Lemma 4 (The proof of the relaxed case starts the same as that of Lemma 3 and bifurcates at the ♠ symbol on page 19)

The first stage of the proof is to show that if at least one of $\mathbf{P}$ or $\mathbf{R}$ has a "relaxed" structure, then both $\mathbf{P}$ and $\mathbf{R}$ must have the same relaxed structure, in the following sense. Applying Lemma 2 to (72) in the relaxed case implies that a non-trivial solution to (72a) exists iff $\mathbf{P}$ has either structure (69) or (68), and similarly for $\mathbf{R}$ (72b). Consider now those non-trivial solutions. According to Lemma 2, all non-zero $\mathbf{M}^{[k]}$ may be full-rank. Consider now some k for which $\mathbf{M}^{[k]}$ is full-rank and assume, w.l.o.g., that $P^{[k]} \geq R^{[k]}$ for this specific k. Then,

$$\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]}\mathbf{L}^{[l]} \implies \mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]}\mathbf{R}^{[k,l]} = \mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]}\mathbf{L}^{[l]} \implies$$
$$\mathbf{R}^{[k,l]} = (\mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]})^{-1}\mathbf{L}^{[k]\dagger}\mathbf{P}^{[k,l]}\mathbf{L}^{[l]} = \begin{cases} \mathbf{0}_{P^{[k]}\times R^{[k]}} & \mathbf{P}^{[k,l]} = \mathbf{0} \\ \neq \mathbf{0} & \mathbf{P}^{[k,l]} \neq \mathbf{0} \end{cases} \tag{76}$$

We conclude that at any k for which $\mathbf{M}^{[k]} \neq \mathbf{0}$, $\mathbf{P}$ and $\mathbf{R}$ must have the same structure (69) or (68), i.e., zeros in the same places. We now discuss non-trivial solutions to (1) for these two structures.

Case C.3: Both $\mathbf{P}$ and $\mathbf{R}$ have structure (68), with possibly different dimensions. The setup $D = K$ has already been dealt with in Cases C.1–C.2. For $2 \leq D < K$, applying the same analysis as in Cases C.1–C.2 to $\mathbf{P}^{[1:D,1:D]}$ and $\mathbf{R}^{[1:D,1:D]}$, we conclude that (1) has non-trivial solutions

$$\mathbf{L}^{[k]} = \begin{cases} \begin{cases} \nu \mathbf{O}^{[k]} & k \in \mathcal{D} \\ \mathbf{0}_{R \times R} & k \notin \mathcal{D} \end{cases} & \text{if } P^{[k]} = R^{[k]} \ \forall k \\ \quad \mathbf{0}_{P \times R} \quad \forall k & \text{if } P^{[k]} \neq R^{[k]} \ \text{ for at least one } k \in \mathcal{D} \end{cases}$$

where $\nu \in \mathbb{C}$ and $\mathbf{O}^{[k]}$ unitary. Note that we set $\mathbf{L}^{[k \notin \mathcal{D}]} = \mathbf{0}$ because this setup does not provide information about what happens in $\mathbf{P}^{[k \notin \mathcal{D}]}$ and $\mathbf{R}^{[k \notin \mathcal{D}]}$.

Case C.4: Both $\mathbf{P}$ and $\mathbf{R}$ have structure (69), with possibly different dimensions. Applying Lemma 2 to (72), for $1 \leq D \leq K$,

$$\mathbf{L}^{[k]}\mathbf{L}^{[k]\dagger} = \mathbf{U}_P^{[k]}\boldsymbol{\Lambda}_P^{[k]}\mathbf{U}_P^{[k]\dagger} \quad , \quad \mathbf{L}^{[k]\dagger}\mathbf{L}^{[k]} = \mathbf{U}_R^{[k]}\boldsymbol{\Lambda}_R^{[k]}\mathbf{U}_R^{[k]\dagger} \quad k \in \mathcal{D} \tag{77}$$

where $\mathbf{U}_P^{[k]}$ and $\mathbf{U}_R^{[k]}$ are the unitary matrices of the EVD of $\mathbf{P}^{[k,k]}$ and $\mathbf{R}^{[k,k]}$, respectively, and $\boldsymbol{\Lambda}_P^{[k]}$ and $\boldsymbol{\Lambda}_R^{[k]}$ arbitrary diagonal matrices. Assume now that for some k, $P^{[k]} \geq R^{[k]}$. By Lemma 2, $\boldsymbol{\Lambda}_R^{[k]}$ is non-zero and may be full-rank. Then, for any $k \in \mathcal{D}$ for which $P^{[k]} \geq R^{[k]}$, (1) implies that $\boldsymbol{\Lambda}_P^{[k]}$ has the same rank as $\boldsymbol{\Lambda}_R^{[k]}$. Denote this rank $Q^{[k]}$. The only option for (1) to hold is that $\mathbf{U}_P^{[k]}$ and $\mathbf{U}_R^{[k]}$ be permutation matrices (any scaling is absorbed by the non-zero values of the diagonal matrices). However, this setup imposes a constraint on $\mathbf{P}$ and $\mathbf{R}$ that forces them to have structure

$$\begin{aligned}
\mathbf{P}^{[k,k]} &= \left[\begin{array}{c|c} \mathrm{diag}\{[\mathbf{P}^{[k,k]}]_{1:Q^{[k]},1:Q^{[k]}}\} & \mathbf{0} \\ \hline \mathbf{0} & [\mathbf{P}^{[k,k]}]_{Q^{[k]}+1:P^{[k]},Q^{[k]}+1:P^{[k]}} \end{array}\right] \\
&= \left[\begin{array}{c|c} \begin{matrix} p_{1,1}^{[k,k]} & & 0 \\ & \ddots & \\ 0 & & p_{Q^{[k]},Q^{[k]}}^{[k,k]} \end{matrix} & \mathbf{0} \\ \hline \mathbf{0}_{(P^{[k]}-Q^{[k]}) \times Q^{[k]}} & [\mathbf{P}^{[k,k]}]_{Q^{[k]}+1:P^{[k]},Q^{[k]}+1:P^{[k]}} \end{array}\right] \\
\mathbf{R}^{[k,k]} &= \left[\begin{array}{c|c} \mathrm{diag}\{[\mathbf{R}^{[k,k]}]_{1:Q^{[k]},1:Q^{[k]}}\} & \mathbf{0} \\ \hline \mathbf{0} & [\mathbf{R}^{[k,k]}]_{Q^{[k]}+1:R^{[k]},Q^{[k]}+1:R^{[k]}} \end{array}\right] \\
&= \left[\begin{array}{c|c} \begin{matrix} r_{1,1}^{[k,k]} & & 0 \\ & \ddots & \\ 0 & & r_{Q^{[k]},Q^{[k]}}^{[k,k]} \end{matrix} & \mathbf{0} \\ \hline \mathbf{0}_{(R^{[k]}-Q^{[k]}) \times Q^{[k]}} & [\mathbf{R}^{[k,k]}]_{Q^{[k]}+1:R^{[k]},Q^{[k]}+1:R^{[k]}} \end{array}\right]
\end{aligned} \tag{78}$$

This structure is admissible in the relaxed sense. We conclude that if for at least one k both $\mathbf{R}$ and $\mathbf{P}$ have a structure (78) then there always exists some non-zero $\mathbf{L}^{[k]}$ for which (1) holds. Lemma 4 follows.

□

D Normalization

The key point for the normalization is that it should take account of the multiset nature of this problem and thus whiten the sources within each dataset, or mixture, separately.

$$\mathbf{P} = \boldsymbol{\Omega}_{ii}\mathbf{S}_{ii}\boldsymbol{\Omega}_{ii}^{\top} \quad \Leftrightarrow \quad \mathbf{S}_{ii} = \boldsymbol{\Omega}_{ii}^{-1}\mathbf{P}\boldsymbol{\Omega}_{ii}^{-\dagger} \tag{79a}$$

$$\mathbf{R} = \boldsymbol{\Omega}_{jj}\mathbf{S}_{jj}\boldsymbol{\Omega}_{jj}^{\top} \quad \Leftrightarrow \quad \mathbf{S}_{jj} = \boldsymbol{\Omega}_{jj}^{-1}\mathbf{R}\boldsymbol{\Omega}_{jj}^{-\dagger} \tag{79b}$$

or in blockwise form,

$$\mathbf{P}^{[k,l]} = \boldsymbol{\Omega}_{ii}^{[k]}\mathbf{S}_{ii}^{[k,l]}\boldsymbol{\Omega}_{ii}^{[l]\top} \quad \in \mathbb{K}^{m_i \times m_i} \tag{79c}$$

$$\mathbf{R}^{[k,l]} = \boldsymbol{\Omega}_{jj}^{[k]}\mathbf{S}_{jj}^{[k,l]}\boldsymbol{\Omega}_{jj}^{[l]\top} \quad \in \mathbb{K}^{m_j \times m_j} \tag{79d}$$

such that

$$\mathbf{P}^{[k,k]} = \mathbf{I}_{m_i} \tag{80a}$$

$$\mathbf{R}^{[k,k]} = \mathbf{I}_{m_j} \, , \tag{80b}$$

where

$$\boldsymbol{\Omega}_{ii}^{[k]} \triangleq (\mathbf{S}_{ii}^{[k,k]})^{-\frac{1}{2}} \quad \in \mathbb{S}_{+}^{m_i \times m_i} \quad , \quad \boldsymbol{\Omega}_{ii} = \bigoplus_{k=1}^{K} \boldsymbol{\Omega}_{ii}^{[k]} \tag{81}$$

E Proof of (15)

Proof of (15). Applying the normalization scheme (79) to (14) leads to

$$\underbrace{\boldsymbol{\Omega}_{ii}^{[k]} \mathbf{M}^{[k]} \boldsymbol{\Omega}_{jj}^{-[k]}}_{\mathbf{L}^{[k]}} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \underbrace{\boldsymbol{\Omega}_{ii}^{-[l]\top} \mathbf{N}^{[l]} \boldsymbol{\Omega}_{jj}^{[l]\top}}_{\mathbf{L}'^{[k]}} \quad \Leftrightarrow \quad \underbrace{\boldsymbol{\Omega}_{ii} \mathbf{M} \boldsymbol{\Omega}_{jj}^{-1}}_{\mathbf{L}} \mathbf{R} = \mathbf{P} \underbrace{\boldsymbol{\Omega}_{ii}^{-\dagger} \mathbf{N} \boldsymbol{\Omega}_{jj}^{\top}}_{\mathbf{L}'} \tag{82}$$

The key point is that due to the normalization, for $l = k$,

$$\underbrace{\boldsymbol{\Omega}_{ii}^{[k]} \mathbf{M}^{[k]} \boldsymbol{\Omega}_{jj}^{-[k]}}_{\mathbf{L}^{[k]}} \underbrace{\mathbf{R}^{[k,k]}}_{\mathbf{I}} = \underbrace{\mathbf{P}^{[k,k]}}_{\mathbf{I}} \underbrace{\boldsymbol{\Omega}_{ii}^{-[k]\top} \mathbf{N}^{[k]} \boldsymbol{\Omega}_{jj}^{[k]\top}}_{\mathbf{L}'^{[k]}} \tag{83}$$

Equation (83) implies that $\mathbf{L}^{[k]} = \mathbf{L}'^{[k]}$ $\forall k$, which concludes the proof. □

References

[1] M. W. Kirson, *Introductory Algebra for Physicists*, 2013, ch. Irreducible representations.

[2] C. W. Curtis, *Pioneers of Representation Theory: Frobenius, Burnside, Schur, and Brauer*, ser. History of Mathematics. Providence, RI, USA: American Mathematical Society, 1999, vol. 15.

[3] A. Alexanderian, "A basic note on group representations and Schurs lemma," http://www4.ncsu.edu/~aalexan3/articles/schur.pdf.

[4] D. Vvedensky, "Group theory course notes," http://www.cmth.ph.ic.ac.uk/people/d.vvedensky/courses.html, 2001.

[5] I. Schur, *Neue Begründung der Theorie der Gruppencharaktere*, ser. Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften zu Berlin, 1905. [Online]. Available: https://books.google.fr/books?id=QoUWSQAACAAJ

[6] D. Lahat, J.-F. Cardoso, and H. Messer, "Identifiability of second-order multidimensional ICA," in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1875–1879.

[7] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms. Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.

[8] D. Lahat and C. Jutten, "Joint blind source separation of multidimensional components: Model and algorithm," in *Proc. EUSIPCO*, Lisbon, Portugal, Sep. 2014, pp. 1417–1421.

[9] ——, "Joint independent subspace analysis using second-order statistics," *IEEE Trans. Signal Process.*, 2016, in press.

[10] ——, "An alternative proof for the identifiability of independent vector analysis using second order statistics," in *Proc. ICASSP*, Shanghai, China, Mar. 2016.

[11] J. Vía, M. Anderson, X.-L. Li, and T. Adalı, "Joint blind source separation from second-order statistics: Necessary and sufficient identifiability conditions," in *Proc. ICASSP*, Prague, Czech Republic, May 2011, pp. 2520–2523.

[12] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, Sep. 2014.

[13] M. Congedo, R. Phlypo, and J. Chatel-Goldman, "Orthogonal and non-orthogonal joint blind source separation in the least-squares sense," in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1885–1889.

[14] J. Chatel-Goldman, M. Congedo, and R. Phlypo, "Joint BSS as a natural analysis framework for EEG-hyperscanning." in *Proc. ICASSP*, Vancouver, Canada, May 2013, pp. 1212–1216.

[15] R. F. Silva, S. Plis, T. Adalı, and V. D. Calhoun, "Multidataset independent subspace analysis extends independent vector analysis," in *Proc. ICIP*, Paris, France, Oct. 2014, pp. 2864–2868.

[16] D. Lahat and C. Jutten, "A generalization to Schur's lemma with an application to joint independent subspace analysis," GIPSA-Lab, Grenoble, France, Tech. Rep. hal-01247899, Dec. 2015.

[17] T. Adalı, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, pp. 18–33, May 2014.

[18] M. Sørensen and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms—part I: Uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 2, pp. 496–522, Apr. 2015.

[19] D. Lahat, T. Adalı, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[20] T. Maehara and K. Murota, "Algorithm for error-controlled simultaneous block-diagonalization of matrices," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 2, pp. 605–620, 2011.

[21] ——, "Error-controlling algorithm for simultaneous block-diagonalization and its application to independent component analysis," *JSIAM Letters*, vol. 2, pp. 131–134, 2010.

[22] ——, "A numerical algorithm for block-diagonal decomposition of matrix *-algebras with general irreducible components," *Japan Journal of Industrial and Applied Mathematics*, vol. 27, no. 2, pp. 263–293, Sep. 2010.

[23] H. W. Gutch, T. Maehara, and F. J. Theis, "Second order subspace analysis and simple decompositions," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, Eds., vol. 6365. Heidelberg: Springer, 2010, pp. 370–377.

[24] D. Lahat and C. Jutten, "On the uniqueness of coupled matrix block diagonalization in the joint analysis of multiple datasets (talk)," in *SIAM Conference on Applied Linear Algebra*, Atlanta, GA, USA, Oct. 2015, http://meetings.siam.org/sess/dsp_talk.cfm?p=72077.

[25] T. W. Anderson, *An introduction to multivariate statistical analysis.* John Wiley & Sons, 1958.

[26] D. Lahat, J.-F. Cardoso, and H. Messer, "Joint block diagonalization algorithms for optimal separation of multidimensional components," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., vol. 7191. Springer, 2012, pp. 155–162.

[27] M. Anderson, "Independent vector analysis: Theory, algorithms, and applications," Ph.D. dissertation, Graduate School of the University of Maryland, Baltimore County, 2013.

[28] J. R. Magnus and H. Neudecker, "The commutation matrix: Some properties and applications," *Ann. Statist.*, vol. 7, no. 2, pp. 381–394, Mar. 1979.

[29] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[30] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms. Part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1022–1032, 2008.

[31] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, ser. Mathematics and its Applications. Chichester, West Sussex, England: Ellis Horwood Limited, 1981.

[32] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Nov. 2012, version 20121115. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274

[33] A. Cichocki, D. Mandic, A. H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[34] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.

[35] S. Liu and G. Trenkler, "Hadamard, Khatri-Rao, Kronecker and other matrix products," *Int. J. Inform. Syst. Sci*, vol. 4, no. 1, pp. 160–177, 2008.

[36] R. Bro, "Multi-way analysis in the food industry: models, algorithms, and applications," Ph.D. dissertation, Chemometrics Group, Food Technology. Department of Dairy and Food Science. Royal Veterinary and Agricultural University, Denmark, 1998.

[37] C. R. Rao and S. K. Mitra, *Generalized inverse of matrices and its applications.* John Wiley & Sons, 1971, vol. 7.

[38] R. P. McDonald, "A simple comprehensive model for the analysis of covariance structures: Some remarks on applications," *Br. J. Math. Stat. Psychol.*, vol. 33, no. 2, pp. 161–183, 1980.

[39] K. Murota, Y. Kanno, M. Kojima, and S. Kojima, "A numerical algorithm for block-diagonal decomposition of matrix *-algebras with application to semidefinite programming," *Japan J. Indust. Appl. Math.*, vol. 27, no. 1, pp. 125–160, Jun. 2010.